

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: January 15, 2003, 15:39:33 ; Search time 7.85318 Seconds  
(without alignments)  
3506.895 Million cell updates/sec

Title: US-09-856-221-4  
Perfect score: 605  
Sequence: 1 gtgaagcgagcattgtcaa.....atttgaaagaaatgaacgg 332

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 225784

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p\_model -DEV=xlp  
-O/cgn2.1/USPTO\_spool/US09856221/runat\_15012003\_153924\_12445/app\_query.fasta\_1.1948  
-DB=SwissProt\_40 -QFWT=fascan -SUFFIX=rsp -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=pet -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09856221 -CGN\_1.1.38 -runat\_15012003\_153924\_12445 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7  
-Ygapop=10 -Ygapext=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	69.5	11.5	1188	1 T1RK_ECOLI	P08956 escherichia
2	69.5	11.5	1437	1 REST_HUMAN	P30622 homo sapien
3	69	11.4	353	1 YKFL_YEAST	P35735 saccharomyc
4	68	11.2	378	1 O3OA_DROME	O9vie5 drosophila
5	65.5	10.8	1353	1 XDH_CALVI	P08793 calliphora
6	65	10.7	211	1 G12_ANOGA	Q17040 anopheles 9
7	65	10.7	320	1 RLPA_RICPR	G9zdel rickettsia
8	65	10.7	534	1 FMO2_CAVPO	P36366 cavia porce
9	65	11.1	620	1 MUTL_TREPA	O83325 treponema p
10	65	10.7	680	1 GAG_SCVLA	P32503 saccharomyc
11	64.5	10.7	286	1 TEHB_HAETN	P45134 haemophilus
12	64.5	10.7	637	1 YHML_YEAST	P38856 saccharomyc
13	64	10.6	919	1 GLK3_HUMAN	Q13003 homo sapien
14	63.5	10.5	223	1 Y836_METJA	G58248 methanococc
15	63.5	10.5	345	1 YDE9_SCHPO	Q10442 schizosacch
16	63.5	10.8	502	1 NUC2_MESVI	Q9muq6 mesostigma
17	63.5	10.5	519	1 LNT_SYNY3	P74055 synechocyst
18	63.5	10.5	1083	1 MAN1_YEAST	P22855 saccharomyc

19	63	10.4	419	1 P47K_PSECL	P31521 pseudomonas
20	63	10.4	423	1 P47A_CANBO	P21245 candida boi
21	63	10.4	465	1 Y065_MYCPN	P75612 mycoplasma
22	62	10.2	419	1 P47B_CANBO	Q00319 candida boi
23	62	10.2	601	1 PEPF_LACLA	Q9cev7 lactococcus
24	61.5	10.2	120	1 ARD2_ECOLI	P52148 escherichia
25	61.5	10.2	153	1 MLJA_XENIA	P51048 xenopus lae
26	61.5	10.2	452	1 NORM_BACSU	O31855 bacillus su
27	61.5	10.2	819	1 AKLH_SERMA	P27725 serratia ma
28	61.5	10.5	1687	1 YDHA_SCHPO	Q92355 schizosacch
29	61.5	10.2	1901	1 YCF1_TORAC	P12222 nicotiana t
30	61	10.1	111	1 TMML_ECOLI	P08701 escherichia
31	61	10.1	176	1 THAP_HUMAN	Q01658 homo sapien
32	61	10.1	292	1 NLA_DROME	Q9xz18 drosophila
33	61	10.1	656	1 YC26_PORPU	P51392 porphyra pu
34	61	10.1	731	1 MASZ_MYCLE	O32913 mycobacteri
35	61	10.1	919	1 GLK3_RAT	P42644 rattus norv
36	60.5	10.3	354	1 MBHS_WOLSU	P31884 wolfinella s
37	60.5	10.0	516	1 YMCA_BACSU	P39599 bacillus su
38	60.5	10.0	547	1 NUSM_ASCSU	P24884 ascaris suu
39	60.5	10.0	913	1 CAD4_MOUSE	P39038 mus musculu
40	60	9.9	309	1 META_YERPE	Q82ar4 yersinia pe
41	60	9.9	329	1 RIR2_BACSU	P50621 bacillus su
42	60	9.9	354	1 DP42_RHIME	Q92xh8 rhizobium m
43	60	9.9	356	1 Y359_AQUAE	O66885 aquifex aeo
44	60	9.9	388	1 TRB1_PYRAB	Q9vlg8 pyrococcus
45	60	9.9	440	1 CYAD_BORPE	P11091 bordetella

## ALIGNMENTS

RESULT 1  
T1RK\_ECOLI  
ID T1RK\_ECOLI STANDARD; PRT: 1188 AA.  
AC P08956;  
DT 01-NOV-1988 (Rel. 09, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Type I restriction enzyme EcoKI R protein (EC 3.1.21.3) (R.EcoKI).  
GN HSDR OR HSR OR B4350.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562;  
[1]  
RN  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12;  
RX MEDLINE=88118919; PubMed=3323532;  
RA Loenen W.A.M., Daniel A.S., Braymer H.D., Murray N.E.;  
RT "Organization and sequence of the hsd genes of Escherichia coli  
X-12";  
RL J. Mol. Biol. 198;159-170(1987).  
[2]  
RN  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / MG1655;  
RX MEDLINE=95334362; PubMed=7610040;  
RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,  
RA Blattner F.R.;  
RT "Analysis of the Escherichia coli genome VI: DNA sequence of the  
region from 92.8 through 100 minutes";  
RL Nucleic Acids Res. 23:2105-2119(1995).  
[3]  
RN  
RP SEQUENCE OF 1-27 FROM N.A.  
RC STRAIN=K12 / CR63;  
RX MEDLINE=91317743; PubMed=1650347;  
RA Waite-Rees P.A., Keating C.J., Moran L.S., Slatko B.E., Hornstra L.J.,  
RA Benner J.S.;  
RT "Characterization and expression of the Escherichia coli Mrr  
restriction system";  
RL J. Bacteriol. 173:5207-5219(1991).  
CC -!- FUNCTION: THE ECOKI ENZYME RECOGNIZES 5'AACN(6)GTGC-3'. SUBUNIT R  
IS REQUIRED FOR BOTH NUCLEASE AND ATPASE ACTIVITIES, BUT NOT FOR  
MODIFICATION.  
CC









```

DR PIR; A29627; A29627.
DR PIR; JQ0407; JQ0407.
DR HSS; S03392; S03392.
DR HSS; P80457; IFO4.
DR InterPro; IPR002888; 2Fe-2S_bind.
DR InterPro; IPR000564; 2Fe2S_ferredoxin.
DR InterPro; IPR000674; AldXan_dh.C.
DR InterPro; IPR005107; CO deh flav.C.
DR InterPro; IPR000572; Euk_Mb_oxred.
DR InterPro; IPR001041; Ferredoxin.
DR InterPro; IPR002346; dehydrog_molyp.
DR Pfam; PF00111; fer2; 1.
DR Pfam; PF00941; FAD_binding_5; 1.
DR Pfam; PF01315; Ald_xan_dh.C; 1.
DR Pfam; PF01799; fer2; 2; 1.
DR Pfam; PF02738; Ald_Xan_dh.C2; 1.
DR Pfam; PF03450; CO deh flav.C; 1.
DR ProDom; PD186071; 2Fe-2S_bind; 1.
DR PROSITE; PS00197; 2FE2S_FERREDOXIN; 1.
DR PROSITE; PS00559; MOLYBDENUM; Flavoprotein; FAD; Iron-sulfur;
KW Oxidoreductase; NAD; Molybdenum; Flavoprotein; FAD; Iron-sulfur;
KW Peroxisome.
FT METAL 50 50 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
FT METAL 56 56 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
FT METAL 61 61 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
FT METAL 64 64 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
SQ SEQUENCE 1353 AA; 150208 MW; 7120361C57C3E297 CRC64;

Alignment Scores:
Pred. No.: 12.9 Length: 1353
Score: 65.50 Matches: 23
Percent Similarity: 44.32% Conservative: 16
Best Local Similarity: 26.14% Mismatches: 28
Query Match: 10.83% Indels: 21
DB: 1 Gaps: 4

US-09-856-221-4 (1-332) x XDP_CALVI (1-1353)
QY 117 GGGCGATTGGCGGCTATTATATACAGTTTATGACTGGCT-----GTT 161
Db 877 GlyArgLeuThrGlyCysTyrIleGluCysTyrAsnAlaGlyTyrSerMetAspLeu 896
QY 162 TCCTGCTGTTGATGGCTGACAAACCTTACCAGTATGAATGACGATAAAGCTGTACGC 221
Db 897 SerPheSerValLeuGluArgAlaMetPheHisPheGlu---AsnCysTyrLysIlePro 915
QY 222 TTCATTAAAGCCGGCTGCTGG----- 242
Db 916 AsnIleLysValGlyGlyTyrValCysLysThrAsnLeuProSerAsnThrAlaPheArg 935
QY 243 ---CATGGCACTATGCTGGTTGTTAGCAGGTGAACCTTGATGCTGAATTTGGCAGAC 299
Db 936 GlyPheGlyGlyProGlnGlyMetPheAlaGlyGluHisIleArgAspValAlaArg 955
QY 300 ---ATGGAAACAACTATTGGAA 320
Db 956 IleLeuGlyLysAspTyrLeuGlu 963

RESULT 6
G12_ANOGA
ID G12_ANOGA STANDARD; PRT; 211 AA.
AC Q17040;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Protein G12 precursor (ANL12).
OS Anopheles gambiae (African malaria mosquito).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
OC Culicoidae; Anopheles.
OX NCBI_TaxID=7165;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN=Suakoko;
RA Mueller H.M., Crisanti A.;
RL Submitted (JUN-1993) to the EMBL/GenBank/DBJ databases.
CC -1- DEVELOPMENTAL STAGE: INDUCED IN THE MIDGUT OF FEMALE AFTER BLOOD MEAL.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z22925; CAA80505.1; -.
KW SIGNAL.
FT CHAIN 1 19 POTENTIAL.
FT CHAIN 20 211 PROTEIN G12.
SQ SEQUENCE 211 AA; 23568 MW; 5BC71P975E5FCF8B CRC64;

Alignment Scores:
Pred. No.: 13.2 Length: 211
Score: 65.00 Matches: 29
Percent Similarity: 45.19% Conservative: 18
Best Local Similarity: 27.88% Mismatches: 36
Query Match: 10.74% Indels: 21
DB: 1 Gaps: 3

US-09-856-221-4 (1-332) x G12_ANOGA (1-211)
QY 68 CTTCTACAAAGCAAAATTCAGACATACAGCTGTGTATACCTGCTACGCTGG---CGATT 124
Db 51 LeuThrAspLysGluValGlnGln-ThrLeuLeuTyr-----LeuGlnGlyGluGluPh 68
QY 125 GCGGCTATTATTATCATCAGTTTATGACTGGCTGTT----- 161
Db 68 eSerAlaValTrpAspGlnPheGluLeuSerAlaValArgAspLeuLeuGlnTyrLe 88
QY 162 -----TCCCTGCTGTTTGTATGCTGAACAACTTACCA 193
Db 88 uGluGluAlaGlyValProAlaTyrGluSerLeuAsnValValAlaAspPheLeuGlyLe 108
QY 194 GTATGAATGACGATTAAGCTGTACGCTTCATTAAGCCGCTGCTGCGATGCACCTTA 253
Db 108 uSerProLeuLysProThrSerValArgSerLeuAlaAlaArgThrGlyGlyLe 128
QY 254 TGTGCTGTTTGTAGCAGGTGAACCTTGTATGCTGAATTTGGCAGACAGATGGAACAACTA 313
Db 128 uAsnGlyLeuLeuGluAlaLeuAlaMetMetProAlaAlaGluLeuAlaMetPh 148
QY 314 TTTGGAAAA 323
Db 148 eGluGluLys 151

RESULT 7
RLPA_RICPR
ID RLPA_RICPR STANDARD; PRT; 320 AA.
AC Q9ZDE1;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE RlpA-like protein.
DE RP390.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=782;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=Madrid E;
RA MEDLINE=99039499; PubMed=9823893;
RA Andersson S.G.E., Zomerdipour A., Andersson J.O.,
RA Scharitz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,

```



OS Treponema pallidum.  
 OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.  
 ON NCBI\_TaxID=160;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Nichols;  
 RX MEDLINE=98332770; PubMed=9665876;  
 RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,  
 RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,  
 RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,  
 RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,  
 RA McDonald L., Artach P., Bowman C., Cotton M.D., Fujii C., Garland S.,  
 RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,  
 RA Venter J.C.;  
 RT "Complete genome sequence of Treponema pallidum, the syphilis  
 spirochete";  
 RL Science 281:375-388(1998).  
 CC -1- FUNCTION: THIS PROTEIN IS INVOLVED IN THE REPAIR OF MISMATCHES IN  
 DNA. IT IS REQUIRED FOR DAM-DEPENDENT METHYL-DIRECTED DNA MISMATCH  
 REPAIR. MAY ACT AS A "MOLECULAR MATCHMAKER", A PROTEIN THAT  
 PROMOTES THE FORMATION OF A STABLE COMPLEX BETWEEN TWO OR MORE  
 DNA-BINDING PROTEINS IN AN ATP-DEPENDENT MANNER WITHOUT ITSELF  
 BEING PART OF A FINAL EFFECTOR COMPLEX (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE DNA MISMATCH REPAIR MUTL/HEXB FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 or send an email to license@isb-sib.ch).  
 CC  
 CC EMBL; AF001210; AAC65291.1; ALT\_INIT.  
 DR HSP; P23367; LBKN.  
 DR TIGR; TP0303; -  
 DR InterPro; IPR003594; ATPbind\_ATPase.  
 DR InterPro; IPR002099; DNA\_mis\_repair.  
 DR Pfam; PF01119; DNA\_mis\_repair; 1.  
 DR Pfam; PF02518; HATPase\_C; 1.  
 DR TIGRams; TIGR00585; multi; 1.  
 DR PROSITE; PS00058; DNA\_MISMATCH\_REPAIR\_1; 1.  
 DR DNA repair; Complete proteome.  
 SQ SEQUENCE 620 AA; 68495 MW; 0604AD802671EA36 CRC64;  
 Alignment Scores:  
 Pred. No.: 14.1 Length: 620  
 Score: 65.00 Matches: 25  
 Percent Similarity: 33.96% Conservative: 11  
 Best Local Similarity: 23.58% Mismatches: 38  
 Query Match: 11.07% Indels: 32  
 DB: 1 Gaps: 5  
 US-09-856-221-4 (1-332) x MUTL\_TREPA (1-620)  
 QY 321 TTTCACAAATGTTTTTTCAC----- 301  
 Db PheProAsnGlyThrPheProValAlaCysLeuPheLeuThrValAsnSerGluArgile 313  
 QY 300 -----TCGTGCGCAATTCACATCAGGTTTCACCTGCTAACAAACAGCATAGTGC 247  
 Db 314 ASPPheAsnIleHisProAlaLysLysGluValHisLeu-----GlnAspTyrAla 330  
 QY 246 CATGCCAGCAGCGGCTTAATGAAGCGTA---CAGCTTTATCGTTTCAATTCATACTGGT 190  
 Db 331 HisIleArgHisThrLeuSerArgSerValAlaHisPheTyrArgGlnCysThrIleAla 350  
 QY 189 AACTTTGTTTCAG-----CCATCAACACAGGAAACAGCCCAAGTCAT 148  
 Db 351 HisTyrValArgAlaGluProAlaHisAlaProAlaThrGlnGlyAsnAlaProThrHis 370  
 QY 147 AAACAT-----GATAATAAATAGCCCAATCCGCCACGTCGCC 109  
 DB: 11  
 TERB\_HAEIN

Db 371 SerSerProProCysThrGlyValArgGluProAlaAlaProCysAlaHisThrPro 390  
 QY 108 AGTTATACAACTGCTGAT 91  
 Db 391 ArgTyrGluSerLeuPhe 396  
 RESULT 10  
 GAG\_SCVLA STANDARD; PRT; 680 AA.  
 ID GAG\_SCVLA  
 AC P32503;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DE 01-OCT-1993 (Rel. 27, Last annotation update)  
 DE Major coat protein.  
 GN GAG.  
 OS Saccharomyces cerevisiae virus L-A (SCV-L-A).  
 OC Viruses; dsRNA viruses; Totiviridae; Totivirus.  
 ON NCBI\_TaxID=11008;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89214077; PubMed=2651431;  
 RA Icho T., Wickner R.B.;  
 RT "The double-stranded RNA genome of yeast virus L-A encodes its own  
 putative RNA polymerase by fusing two open reading frames";  
 RL J. Biol. Chem. 264:6716-6723(1989).  
 CC -1- PTM: ACETYLATION IS NECESSARY FOR VIRAL ASSEMBLY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 or send an email to license@isb-sib.ch).  
 CC  
 CC EMBL; J04592; AAA50506.1; -  
 DR PIR; S26764; S26764.  
 KW Coat protein; Acetylation.  
 FT MOD\_RES 1  
 SQ SEQUENCE 680 AA; 75994 MW; 32353D8B7F05C943 CRC64;  
 Alignment Scores:  
 Pred. No.: 14.2 Length: 680  
 Score: 65.00 Matches: 21  
 Percent Similarity: 41.30% Conservative: 17  
 Best Local Similarity: 22.83% Mismatches: 36  
 Query Match: 10.74% Indels: 18  
 DB: 1 Gaps: 2  
 US-09-856-221-4 (1-332) x GAG\_SCVLA (1-680)  
 QY 51 ACTCAGGCACAGCTGGCGCTTCCTCAACAAAGCAATTCAGCAATACAGCGTTGTATACTGG 110  
 Db 269 SerLysValMetLeuSerAlaLeuArgLysTyrValAsnHisAsnArgLeuTyrAsn--- 287  
 QY 111 CTACGTGGCGCATTTGGCGGCTATTATTATATACAGTTTATGACTTGGCTGTTCCCTCGT 170  
 Db 288 -----GlnPheTyrThrAlaAlaGlnLeuLeuAla 297  
 QY 171 TTGATG-----GCTGAACAAACTTACCAGTATGATGAATGAAGCAT 209  
 Db 298 GlnIleMetMetLysProValProAsnCysAlaGluGlyTyrAlaTyrLeuMetHisAsp 317  
 QY 210 AAAGCTGTACCTTCATTAAAGCCCGGTGCTGCATGCGCATTATGCTGGTTTGTAGCA 269  
 Db 318 AlaLeuValAsnIleProLysPheGlySerIleArgGlyArgTyrProPheLeuLeuSer 337  
 QY 270 GGTGAACACTTGTAGCTGAATTTGGCACAGATGAA 305  
 Db 338 GlyAspAlaAlaLeuIleGlnAlaThrAlaLeuGlu 349  
 RESULT 11  
 TERB\_HAEIN

```

ID TEHB_HAEIN STANDARD; PRT; 286 AA.
AC P45134;
DT 01-NOV-1995 (Rel. 32, Created)
DE 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tellurite resistance protein tehB homolog.
GN TEHB OR H1275.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A. / ATCC 51907;
RC MEDLINE=9530630; PubMed=7542800;
RX Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA Kenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Otterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RD ";
RL Science 269:496-512(1995).
CC -!- FUNCTION: INVOLVED IN POTASSIUM TELLURITE RESISTANCE (POTENTIAL).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U32807; AAC22923.1; -.
DR TIGR: H11275;
DR InterPro: IPR000051; SAM_bind.
DR InterPro: IPR004537; TehB.
DR TIGRFAMs: TIGR00477; tehB. 1.
KW Antibiotic resistance; Tellurium resistance; Complete proteome.
SQ SEQUENCE 286 AA; 33019 MW; 69D7647E97CC74FA CRC64;

Alignment Scores:
Pred. No.: 15.4 Length: 286
Score: 64.50 Matches: 17
Percent Similarity: 52.63% Conservative: 13
Best Local Similarity: 29.82% Mismatches: 14
Query Match: 10.66% Indels: 13
DB: 1 Gaps: 3

US-09-856-221-4 (1-332) x TEHB_HAEIN (1-286)
QY 90 AATACAGCGTTGTATAACTGG-----CTACGTGGCGGATGGCGGTATT 134
DB AenThrLysValGlyThrTrpGlyLysLeuThrValLeuLysGlyLysLeu----- 44
QY 135 TATTATCATGTTTATGACTTGCTGTTCCCTGCTGTGATGGCTGAACAACTTACCAG 194
DB 45 -----LysPheTyrgluLeuThrGluAsnGlyAspValIleAlaGluHisIlePheThr 62
QY 195 TATGAATGAACATAAGCTTTCATTAAGCCCGGCTGGCAT 245
DB 63 ProGluSerHis-----IleProPheValGluProGlnAlaTrpHis 76

RESULT 12
YHWM_YEAST STANDARD; PRT; 637 AA.
ID YHWM_YEAST
AC P38856;
DT 01-FEB-1995 (Rel. 31, Created)

```

```

DT 01-FEB-1995 (Rel. 31, Last sequence update)
DE 01-OCT-1996 (Rel. 34, Last annotation update)
DE Hypothetical 71.7 kDa protein in REC104-SOL3 intergenic region.
GN YHR161C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=94378003; PubMed=8091229;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
Du Z., Favell A., Fulton L., Gattung S., Geisel C., Kirsten J.,
RA Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,
RA Latreille P., Louis E.J., Macri C., Mardis E., Menezes S., Mouser L.,
RA Nham M., Rifkin L., Riles L., St Peter H., Trevaskis E., Vaughan K.,
RA Vignati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,
RA Vaudin M.;
RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
RT VIII."
RL Science 265:2077-2082(1994).
CC -!- SIMILARITY: TO YEAST YGR241C.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U10397; AAF68993.1; -.
DR PIR: S46771; S46771.
DR SGD: S0001204; YAP1801.
DR InterPro: IPR001026; ENTH.
DR Pfam: PF01417; ENTH; 1.
DR SMART: SM00273; ENTH; 1.
DR Hypothetical protein.
FT DOMAIN 560 579 POLY-GLN.
SQ SEQUENCE 637 AA; 71660 MW; 76B36BELL1697A8C2 CRC64;

Alignment Scores:
Pred. No.: 16.2 Length: 637
Score: 64.50 Matches: 28
Percent Similarity: 43.09% Conservative: 25
Best Local Similarity: 22.76% Mismatches: 45
Query Match: 10.66% Indels: 25
DB: 1 Gaps: 5

US-09-856-221-4 (1-332) x YHWM_YEAST (1-637)
QY 24 AACTATCTGGAACCCAACTGACGACAGCAGCTG-----GCCTTCCTACAA 77
DB 156 AsPHisValGluSerLeuGluValGlnIleGlnAlaLeuLysAsnLysTyThrGln 175
QY 78 AGCAATTCAGCAAT-----ACA 95
DB 176 TyrAspLeuSerAsnGluLeuIleIlePheGlyPheLysLeuLeuIleGlnAspLeuLeu 195
QY 96 GCCTTCGTAACCTGACGTGGCGGATGGCGGCTATTATATATATATATATATATATATATAT 155
DB 196 AlaLeuTyAsnAlaLeuAsnGluGlyIleIleThrLeuLeuGluSerPhePheGluLeu 215
QY 156 GCTGTTTCCTCTGTTTGTGATGGCTGAACAACTTACCAG-----TATGAA 200
DB 216 SerHisHis-----AsnAlaGluArgThrLeuAspLeuTyThrPheValAsp 232
QY 201 TTCAACGATAAGCTGTACGCTTCATTAAAGCCCGGCTGGCGGCTATCTATCTATCTATCT 260
DB 233 LeuThrGluHisValArgTyLeuLysSerGlyLysThrAlaGlyLeuLysIlePro 252
QY 261 TTGTTAGCAGGTGAACCTTGATGCTGAATTTGGCACAGATGGAAGAAAAAAGCTATTGTGAA 320

```

```

Db 253 VallileuylHisIleThrThrylsteu---ValArgSerLeuGluGluHisLeuIleGlu 271
QY 321 AAGATGAA 329
Db 272 AspAspLys 274
RESULT 13
GLK3_HUMAN
ID GLK3_HUMAN STANDARD; PRT; 919 AA.
AC Q13003; Q16136; Q13004;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Glutamate receptor, ionotropic kainate 3 precursor (Glutamate receptor
DE 7) (GluR-7) (GluR7) (Excitatory amino acid receptor 5) (EAA5).
GN GRK3 OR GLUR7.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND VARIANTS SER-310 AND ARG-352.
RC TISSUE=Fetal brain;
RX MEDLINE=95236038; PubMed=7719709;
RA Nutt S.L., Hoo K.H., Rampersad V., Deverill R.M., Elliott C.E.,
RA Fletcher E.J., Adams S.-L., Korczak B., Foides R.L., Kamboj R.K.;
RT "Molecular characterization of the human EAA5 (GluR7) receptor: a
RT high-affinity kainate receptor with novel potential RNA editing
RT sites.";
RL Recept. Channels 2:315-326(1994).
[2]
RN [2]
RP SEQUENCE OF 268-320 FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=94174381; PubMed=8128318;
RA Puranam R.S., Eubanks J.H., Heinemann S.F., McNamara J.O.;
RT "Chromosomal localization of gene for human glutamate receptor
RT subunit-7.";
RL Somat. Cell Mol. Genet. 19:581-588(1993).
[3]
RN [3]
RP VARIANT SER-310.
RX MEDLINE=20574843; PubMed=11124978;
RA Schiffer H.H., Swanson G.T., Masliash E., Heinemann S.F.;
RT "Unequal expression of allelic kainate receptor GluR7 mRNAs in human
RT brains.";
RL J. Neurosci. 20:9025-9033(2000).
CC -!- FUNCTION: L-GLUTAMATE ACTS AS AN EXCITATORY NEUROTRANSMITTER AT
CC MANY SYNAPSES IN THE CENTRAL NERVOUS SYSTEM. THE POSTSYNAPTIC
CC ACTIONS OF GLU ARE MEDIATED BY A VARIETY OF RECEPTORS THAT ARE
CC NAMED ACCORDING TO THEIR SELECTIVE AGONISTS. THIS RECEPTOR BINDS
CC DOMOATE > KAINATE >> L-GLUTAMATE = QUISQUALATE >> AMPA = NMDA.
CC -!- SUBUNIT: HOMOMERIC. CAN ALSO COASSEMBLE WITH EITHER GRIK4 OR
CC GRIK5 TO FORM HETEROMERIC RECEPTORS.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: BELONGS TO THE LIGAND-GATED IONIC CHANNEL FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; U16127; AAB60407.1; -.
CC EMBL; U16128; AAC50421.1; -.
CC EMBL; S69349; AAB30157.1; -.
CC HSP; P19491; IGR2.
CC Genew; HGNC:4581; GRIK3.
CC MIN; I38243; -.
CC InterPro; IPR001828; ANF_receptor.
CC InterPro; IPR001320; Ion_glu_receptor.
CC InterPro; IPR001311; SBP/glu_receptor.

```

```

DR Pfam; PF00060; lig_chan; 1.
DR Pfam; PF01094; ANF_receptor; 1.
DR ProDom; PD000500; Ion_glu_receptor; 1.
DR SMART; SM00079; PBPE; 1.
KW Receptor; Postsynaptic membrane; Ionic channel; Glycoprotein; Signal;
KW Transmembrane; RNA editing; Polymorphism.
FT SIGNAL 1 31 POTENTIAL.
FT CHAIN 32 919
FT DOMAIN 32 563 GLUTAMATE RECEPTOR, IONOTROPIC KAINATE 3.
FT TRANSMEM 564 584 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 637 657 POTENTIAL.
FT TRANSMEM 821 841 POTENTIAL.
FT CARBOHYD 70 70 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 76 76 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 278 278 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 381 381 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 415 415 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 426 426 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 433 433 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 548 548 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARIANT 310 310 A -> S.
FT VARIANT 352 352 /FTid-VAR_000308.
FT VARIANT 303 303 Q -> R (IN RNA EDITED VERSION).
FT CONFLICT 303 303 /FTid-VAR_000309.
FT CONFLICT 303 303 R -> L (IN REF. 2).
SQ SEQUENCE 919 AA; 104031 MW; 2CD7E49B00195E68 CRC64;
Alignment Scores:
Pred. No.: 18.9 Length: 919
Score: 64.00 Matches: 22
Percent Similarity: 40.18% Conservative: 23
Best Local Similarity: 19.64% Mismatches: 61
Query Match: 10.58% Indels: 6
DB: 1 Gaps: 1
US-09-856-221-4 (1-332) x GLK3_HUMAN (1-919)
QY 6 GCGGCGAGTATTGCAAAAAAATACTATCTGGAACCAACAGCAGCAAACTCAGGCACAGCTG 65
Db 293 SerAlaIleValGluLysTrpSerMetGluArgLeuGlnAlaAlaProArgAlaGluSer 312
QY 66 GCCTTCCTACAAAGCAAAATTCAGCAATACAGCGTTGTATTAAGTGGTACGTGGCGGATTTG 125
Db 313 GlyLeuLeuAspGlyValMetMetThrAspAlaAlaLeuLeuTyTrpAspAlaValHisIle 332
QY 126 GCGGCTATTATTATTCAGTTTTATGACTTGGCTGTTCCTGCTGTTTGTATGGCTCAACAA 185
Db 333 ValSerValCysTyTrpGlnArgAlaProGlnMetThrValAsnSerLeuGlnCysHisGln 352
QY 186 ACTTACCACTATGAATTGAACGATAAAGCTGCTAGCTTATTAAAGCCCGGTCCCTGGCAT 245
Db 353 HisLysAlaIleTrpArgPheGlyGlyArgPheMetAsnPhelIleLysGluAlaGlnTrpGlu 372
QY 246 GGCACCTTATGCTGGTTTGTAGCAGGTGAACCC-----TTGATGCTG 287
Db 373 GlyLeuThrGlyArgIleValPheAsnLysThrSerGlyLeuArgThrAspPheAspLeu 392
QY 288 AATTGGGCACATGGAATAAACTATTGGAAAAA 323
Db 393 AsplIleSerLeuLysGluAspGlyLeuGluLys 404
RESULT 14
Y838_METJA
ID Y838_METJA STANDARD; PRT; 223 AA.
AC Q58248;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MJ0838.
GN MJ0838.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.

```

```

OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337993; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts R.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii."
RL Science 273:1058-1073(1996).
CC -!- SIMILARITY: TO M.JANNASCHII MJ0575.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U67528; AAB98843.1; -
DR TIGR; MJ0838; -
DR InterPro: IPR003745; DUF166.
DR Pfam: PF02593; DUF166; 1
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 223 AA; 25837 MW; 1F592F76C84EC37F CRC64;

Alignment Scores:
Pred. No.: 19.9 Length: 223
Score: 63.50 Matches: 16
Percent Similarity: 56.41% Conservative: 6
Best Local Similarity: 41.03% Mismatches: 10
Query Match: 10.50% Indels: 7
DB: 1 Gaps: 2

US-09-856-221-4 (1-332) x Y838_METJA (1-223)
QY 147 TATGACTGGCTGTTCCCTGCTTGTGCTGCTGAACAACTTACCAGTAT----- 197
Db 52 TyrAspLeuPheIleThrTyrThrLeuAsnProAspLeuThrTyrGluLeuValArgLys 71
QY 198 -----GAATTGAACGATGAAGCTGTACGCTTCATTAAAGCCGGTGCCTGGCATGGC 248
Db 72 IleLysGluLeuAsnAsnLysAla-----PheValLeuValGlyAlaTrpLysGly 88

RESULT 15
YDE9_SCHPO
ID YDE9_SCHPO STANDARD; PRT; 345 AA.
AC Q10442;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Putative mitochondrial carrier Cl2B10.09.
OS SPAC12B10.09.
SN Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,

```

```

RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Motlier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerretti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe."
RL Nature 415:871-880(2002).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial
CC inner membrane (Potential).
CC -!- DOMAIN: COMPOSED OF THREE HOMOLOGOUS DOMAINS.
CC -!- SIMILARITY: BELONGS TO THE MITOCHONDRIAL CARRIER FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z70721; CAA94699.1; -
DR InterPro: IPR001993; Mitoch_carrier.
DR Pfam: PF00153; mito_carr; 3
DR PROSITE; PS00215; MITOCH_CARRIER; 2.
DR KW Hypothetical protein; Mitochondrion; Inner membrane; Repeat;
DR Transmembrane; Transport.
DR TRANSMEM 1 21 POTENTIAL.
FT TRANSMEM 83 103 POTENTIAL.
FT TRANSMEM 128 148 POTENTIAL.
FT TRANSMEM 220 240 POTENTIAL.
FT TRANSMEM 262 282 POTENTIAL.
FT TRANSMEM 296 316 POTENTIAL.
FT TRANSMEM 319 339 POTENTIAL.
SQ SEQUENCE 345 AA; 38451 MW; 39081A6AAC984B2C CRC64;

Alignment Scores:
Pred. No.: 20.4 Length: 345
Score: 63.50 Matches: 28
Percent Similarity: 46.84% Conservative: 9
Best Local Similarity: 35.44% Mismatches: 29
Query Match: 10.50% Indels: 13
DB: 1 Gaps: 4

US-09-856-221-4 (1-332) x YDE9_SCHPO (1-345)
QY 40 AACAGGCACAAACTCAGG-----CACAGCTGGCGCTTCCTACAAAGCAATTCAGCA 90
Db 61 AsnArgHisGlnLeuLeuArgThrCysHisPheThrProSer---LysArgHisSerAla 79
QY 91 ATACAGCGTTGTATACCTGCGCTACGTGGCGGATGGCGCTATTTATTCAGTTTATG 150
Db 80 Met--SerPhePheGluAlaLeuGlyAlaGlyIleCysAla-----G 93
QY 151 ACTTGTGCTTCCCTGCTGTTTGTATGGCTGCTGAACAACTTACCAGTATGAATTCAGCAT 210
Db 93 IlyeAlaValAspLeuSerLeuPheProIleAspThrLeuLysThrArgLeuGlnAlaL 113
QY 211 AAGCTGTACGCTTCATTAAAGCCCGCGTGCCTGGCATGCGCACTTATGCTGTTTG 263

```

Db 113 ysGly---GlyPheValLysAsnGlyGlyPheHisGlyValTyrArgGlyLeu 129

Search completed: January 15, 2003, 15:46:35  
Job time : 11.8532 secs





GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: January 15, 2003, 15:44:13 ; Search time 31.2773 Seconds  
(without alignments)  
4374.264 Million cell updates/sec

Title: US-09-856-221-4  
Perfect score: 605  
Sequence: 1 gtaagcgcagctattgca.....atttgaaaaagatgaacgg 332

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 671580 seqs, 206047115 residues  
Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+n2p.model -DRV=rlp  
-Q/cgn2\_1/USPTO.spool/US09856221/runat\_15012003.153925\_12457/app-query.fasta\_1.1948  
-DB=SPTREMBL\_21 -OPMT=fastan -SUFFIX=rspt -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09856221@cgn.1.171.runat\_15012003.153925\_12457 -NCPU=6 -ICPU=3  
-NO\_XUPXY -NO\_MMAPP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SPTREMBL\_21.\*  
1: sp.archaea.\*  
2: sp.bacteria.\*  
3: sp.fungi.\*  
4: sp.human.\*  
5: sp.invertebrate.\*  
6: sp.mammal.\*  
7: sp.mhc.\*  
8: sp.organelle.\*  
9: sp.phage.\*  
10: sp.plant.\*  
11: sp.rodent.\*  
12: sp.virus.\*  
13: sp.vertebrate.\*  
14: sp.unclassified.\*  
15: sp.rviro.\*  
16: sp.bacteriaph.\*  
17: sp.archaea.\*

pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	534	88.3	2523	2 Q93RP0	Q93rp0 xenorhabdus

2	446	73.7	2516	2 Q9RN43	Q9rn43 photorhabdu
3	418	69.1	2504	2 Q85160	Q85160 photorhabdu
4	395.5	65.4	2376	2 Q9F923	Q9f923 serratia en
5	373	61.7	2538	2 Q93RN7	Q93rn7 xenorhabdus
6	268	44.3	1565	2 Q85156	Q85156 photorhabdu
7	219	36.2	1279	2 Q93RP4	Q93rp4 xenorhabdus
8	179.5	29.7	1189	2 Q85152	Q85152 photorhabdu
9	76	12.6	121	2 P96782	P96782 haemophilus
10	75	12.4	188	7 Q8SNH2	Q8snh2 xiphophorus
11	74.5	12.3	598	2 Q9L815	Q9l815 treponema p
12	73	12.0	439	17 Q8431	Q8431 pyrococcus
13	72.5	12.0	1606	3 Q9UW87	Q9uw87 candida alb
14	72	11.9	269	16 Q69492	Q69492 mycobacteri
15	71	11.7	332	2 Q9F7M4	Q9f7m4 uncultured
16	70.5	11.7	553	3 Q8X0K4	Q8x0k4 neorospira
17	70.5	11.7	596	2 Q9K2N7	Q9k2n7 treponema p
18	70.5	11.7	598	16 Q88138	Q88138 treponema p
19	70.5	11.7	1309	16 Q9KS77	Q9ks77 vibrio chol
20	70	11.6	725	4 Q96MR5	Q96mr5 homo sapien
21	69.5	11.5	427	2 Q8VUK1	Q8vuk1 paracoccus
22	69.5	11.5	467	16 Q9JRH0	Q9jrh0 neisseria m
23	69.5	11.5	601	16 Q99253	Q99253 streptococc
24	68.5	11.3	258	16 Q55454	Q55454 synchocyst
25	68.5	11.3	312	5 Q61892	Q61892 caenorhabdi
26	68.5	11.3	363	17 Q97LK7	Q97lk7 sulfolobus
27	68	11.2	503	16 Q8ZG18	Q8zgi8 yersinia pe
28	68	11.2	582	5 Q9N515	Q9n515 caenorhabdi
29	68	11.2	698	16 Q99UJ8	Q99uj8 staphylococ
30	68	11.2	1056	3 Q9P8P8	Q9p8p8 candida alb
31	67.5	11.2	120	2 Q9KJ14	Q9kji4 klebsiella
32	67.5	11.5	335	8 Q03370	Q03370 procavia ca
33	67.5	11.2	596	2 Q9L814	Q9l814 treponema p
34	67.5	11.2	610	5 Q20798	Q20798 caenorhabdi
35	67.5	11.2	680	12 Q87024	Q87024 saccharomyc
36	67.5	11.2	703	2 Q9R714	Q9r714 helicobacte
37	67.5	11.2	703	16 Q92JY3	Q92jy3 helicobacte
38	67.5	11.2	707	2 Q51811	Q51811 helicobacte
39	67.5	11.2	716	5 Q9N8E5	Q9n8e5 trypanosoma
40	67.5	11.2	741	2 Q52269	Q52269 helicobacte
41	67.5	11.2	744	16 Q9ZK2	Q9zk2 helicobacte
42	67	11.1	105	2 P96780	P96780 haemophilus
43	67	11.4	335	8 Q9ZY32	Q9zy32 orycteropus
44	67	11.4	379	8 Q9THD6	Q9thd6 orycteropus
45	67	11.1	590	10 Q9SLW2	Q9slw2 olea europa

## ALIGNMENTS

RESULT 1	Q93RP0	PRELIMINARY;	PRT; 2523 AA.
ID	Q93RP0		
AC	Q93RP0;		
DT	01-DEC-2001 (TREMBlrel. 19, Created)		
DT	01-DEC-2001 (TREMBlrel. 19, Last sequence update)		
DT	01-MAR-2002 (TREMBlrel. 20, Last annotation update)		
DE	XptA1 protein.		
CN	XPTA1		
OS	Xenorhabdus nematophilus.		
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;		
OC	Xenorhabdus.		
OX	NCBI_TaxID=628;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=PMF1296;		
RA	MEDLINE=21218513; PubMed=11319082;		
RA	Morgan J.A.W., Sergeant M., Ellis D., Ousley M., Jarrett P.;		
RT	"Sequence analysis of insecticidal genes from Xenorhabdus nematophilus		
RT	PMF1296.";		
RL	Appl. Environ. Microbiol. 67:2062-2069(2001).		
DR	EMBL; AJ308438; CAC38401.1; -		
DR	InterPro; IPR000953; Chromo.		
DR	SMART; SM00298; CHROMO; 1.		
SQ	SEQUENCE 2523 AA; 286999 MW; 3159852E0655B5B1 CRC64;		





```

O85156
ID O85156 PRELIMINARY; PRT; 1565 AA.
AC O85156;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Insecticidal toxin complex protein TccB.
GN TCCB.
OS Photobacterium luminescens (Xenorhabdus luminescens).
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Photobacterium.
OX NCBI_TaxID=29488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=W-14;
RA Bowen D., Rocheleau T.A., Blackburn M., Andreev O., Golubeva E.,
RT Bhartiya R., french-Constant R.H.;
RA "Insecticidal toxins from the bacterium Photobacterium luminescens.";
RL Science 280:2129-2132(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=W14;
RX MEDLINE=21185117; PubMed=11286884;
RA Waterfield N.R., Bowen D.J., Fetherston J.D., Perry R.D.,
RA french-Constant R.H.;
RT "The tc genes of Photobacterium: a growing family.";
RL Trends Microbiol. 9:185-191(2001).
DR EMBL: AF047028; AAC38629.1; -.
DR EMBL: AF346499; AAL18472.1; -.
SQ SEQUENCE 1565 AA; 175717 MW; 66C3AB96C5FA8397 CRC64;

Alignment Scores:
Pred. No.: 1.5e-23 Length: 1565
Score: 268.00 Matches: 52
Percent Similarity: 61.82% Conservative: 16
Best Local Similarity: 47.27% Mismatches: 42
Query Match: 44.30% Indels: 0
DB: 2 Gaps: 0

US-09-856-221-4 (1-332) x O85156 (1-1565)
QY 3 GAAGCGGAGTATTGCAAAAAAACTCTGGAAACCAACAGGACAACTCAGGCACAG 62
Db LysAlaAlaGlnThrSerLeuGlnGlnAlaLysAlaGlnGlnValGlnIleArgThrMet 1273
QY 63 CTGGCGTCTCTACAAAGCAATTCAGCATACAGCGTGTGTATACCTGCTACGTGGCGGA 122
Db LeuThrTyrLeuThrThrArgPheThrGlnAlaThrLeuTyrGlnThrLeuSerGlyGln 1293
QY 123 TTGGCGGCTATTATTATCATCTTTTATGACTTGGCTGGCTGTTCCTGTTGTTGATGGCTGAA 182
Db LeuSerAlaLeuTyrTyrGlnAlaTyrAspAlaValAlaLeuCysLeuSerAlaGln 1313
QY 183 CAACCTACCATGATGATGACGATGAAGCTGACGCTTCATTAAGCCGGTGCCTGG 242
Db AlaCysTrpGlnTyrGluLeuGlyAspTyrAlaThrPheIleGlnThrGlyThrTrp 1333
QY 243 CATGGCACTTATGCTGGTGTGTAGCAGGTGAAACCTTGATGCTGAATTTGGCAGCATG 302
Db AsnAspHisTyrArgGlyLeuGlnValGlyGluThrLeuGlnLeuAsnLeuHisGlnMet 1353
QY 303 GAAAAAACTATTGGAAAAAGATGAACGG 332
Db GluAlaAlaTyrLeuValAlaArgHisGluArg 1363

RESULT 7
Q93RP4
ID Q93RP4 PRELIMINARY; PRT; 1279 AA.
AC Q93RP4;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)

DE XptD1 protein (Fragment).
GN XPTD1.
OS Xenorhabdus nematophilus
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Xenorhabdus.
OX NCBI_TaxID=628;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=21218513; PubMed=11319082;
RA Morgan J.A.W., Sergeant M., Ellis D., Ousley M., Jarrett P.;
RT "Sequence analysis of insecticidal genes from Xenorhabdus nematophilus
RT PMF1296.";
RL Appl. Environ. Microbiol. 67:2062-2069(2001).
DR EMBL: AJ308438; CAC38397.1; -.
DR InterPro: IPR001993; Mitoch_Carrier.
DR PROSITE: PS00215; MITOCH_CARRIER; UNKNOWN_1.
FT NON_TER 1
SQ SEQUENCE 1279 AA; 146105 MW; 11841387EF906B98 CRC64;

Alignment Scores:
Pred. No.: 1.24e-17 Length: 1279
Score: 219.00 Matches: 41
Percent Similarity: 61.39% Conservative: 21
Best Local Similarity: 40.59% Mismatches: 39
Query Match: 36.20% Indels: 0
DB: 2 Gaps: 0

US-09-856-221-4 (1-332) x Q93RP4 (1-1279)
QY 30 CTGGAACCCCAACAGGACAACTCAGGCACAGCTGGCGCTTCCTACAAACAAATTCAGC 89
Db LeuGluAlaGlnHisGluGlnAspGlnValLeuLeuGluTyrTyrSerAsnArgPheThr 999
QY 90 AATACAGCGTTGTATACCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 149
Db AsnAspAlaLeuTyrMetTyrMetIleSerGlnIleSerGlyLeuTyrLeuGlnAlaTyr 1019
QY 150 GACTTGGCTGTTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 209
Db AspAlaValAsnSerLeuCysLeuLeuAlaGluAlaSerTrpGlnTyrGluThrGlyGln 1039
QY 210 AAGCTGTGACCTCTCATTAAGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 269
Db TyrAspMetAsnPheValGlnSerGlyLeuTyrPAsnAspLeuTyrGlnGlyLeuLeuVal 1059
QY 270 GGTGAACCTTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 329
Db GlyGluHisLeuTyrLeuAlaLeuGlnArgMetAspGlnAlaTyrLeuGlnHisAsnThr 1079
QY 330 CGG 332
Db 1080 Arg 1080

RESULT 8
O85152
ID O85152 PRELIMINARY; PRT; 1189 AA.
AC O85152;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Insecticidal toxin complex protein TccB.
GN TCCB.
OS Photobacterium luminescens (Xenorhabdus luminescens).
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Photobacterium.
OX NCBI_TaxID=29488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=W-14;
RX MEDLINE=98306208; PubMed=9641921;
RA Bowen D., Rocheleau T.A., Blackburn M., Andreev O., Golubeva E.,
RA Bhartiya R., french-Constant R.H.;

```

```

RT *Insecticidal toxins from the bacterium Phototribadus luminescens.";
RL Science 280:2129-2132(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-W14;
RX MEDLINE-21185117; PubMed-11286884;
RA Waterfield N.R., Bowen D.J., Fetherston J.D., Perry R.D.,
  French-Constant R.H.;
RT "The tc genes of Phototribadus: a growing family.";
RL Trends Microbiol. 9:185-191(2001).
DR EMBL; AF046867; AAC38624.1; -.
DR EMBL; AF346497; AAL18450.1; -.
DR InterPro; IPR002927; Virus_Hs.
DR Pfam; PF01550; Virus_Hs; 1.
SQ SEQUENCE 1189 AA; 131466 MW; 7DAA3AE0BAB88127 CRC64;

Alignment Scores:
Pred. No.: 7,34e-13 Length: 1189
Score: 179.50 Matches: 37
Percent Similarity: 58.16% Conservative: 20
Best Local Similarity: 37.76% Mismatches: 40
Query Match: 29.67% Indels: 1
DB: 2 Gaps: 1

US-09-856-221-4 (1-332) x 085152 (1-1189)

QY 33 GAACCCACAGCAGCAAACTCAGCAGCTGGCTTCCTACAAAGCAAAATTCAGCAAT 92
Db 921 GLUThrGluGlnAlaAsnAlaGlnAlaLeuTyrAspLeuGlnThrThrArgPheThrGly 940
QY 93 ACAGCGTGTGTAACCTGCTACCTGGCGGCTATTATTTATATCAGTTTATGAC 152
Db 941 GlnAlaLeuTyrAsnTrpMetAlaGlyArgLeuSerAlaLeuTyrTyrGlnMetTyrAsp 960
QY 153 TTGGCTGTTCCCTGCTTTGATGGCTGAGCAAACTTACCAGTAGTATGATTAACAGATAA 212
Db 961 SerThrLeuProIleCysLeuGlnAlaLysAlaAlaLeuValGlnGluLeuGlyGluLys 980
QY 213 --GCTGTACGCTTCATTAAAGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 269
Db 981 GluSerAspSerLeuPheGlnValProValTrpAsnAspLeuTrpGlnGlyLeuLeuAla 1000
QY 270 GGTGAACCTGCTGCTGAATTTGGCAGCAGATGGAAAGAACTATTTGGAAAGA 323
Db 1001 GlyGluGlyLeuSerSerGluLeuGlnLysLeuAspAlaLeuTrpLeuAlaArg 1018

RESULT 9
P96782 ID P96782 PRELIMINARY; PRT; 121 AA.
AC P96782;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DE Major outer membrane protein P2-type 7 (Fragment).
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RA Smith-Vaughan H.C., Sriprakash K.S., Mathews J.D., Kemp D.J.;
RT "Non-encapsulated Haemophilus influenzae in Aboriginal infants with
  otitis media: Prolonged carriage of P2 porin variants and evidence for
  horizontal P2 gene transfer.";
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; U89638; AAB49486.1; -.
DR InterPro; IPR001702; Porin_gram-ve.
DR Pfam; PF00267; gram-ve_porins; 1.
DR NON_TER 1
FT NON_TER 121
SQ SEQUENCE 121 AA; 13482 MW; D0C7B580FC7F9766 CRC64;

Alignment Scores:
Pred. No.: 7,34e-13 Length: 121
Score: 179.50 Matches: 37
Percent Similarity: 58.16% Conservative: 20
Best Local Similarity: 37.76% Mismatches: 40
Query Match: 29.67% Indels: 1
DB: 2 Gaps: 1

```

```

US-09-856-221-4 (1-332) x P96782 (1-121)

QY 24 AACATATCTGGAACCCACAGCAGCAAACTCAGCAGCTGGCTTCCTACAAAGCAAA 83
Db 55 AsnTyrLysAspThrAsnGlnGlnGlnGln-----GlnLysLys 68
QY 84 TTCAGCAATACAGCGTTGTATTAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 143
Db 69 PheLeuTyrGluGlyLysGlnGluValAsnGlyAlaLeuAlaSerLeuGlyTyrArg 88
QY 144 TTTTATGACTGGCT-----GTTTCCCTGCTGTTTGTATGCTGCTGCTGCTGCTGCTGCTGCT 197
Db 89 PheSerAspLeuGlyLeuLeuValSerLeuAspSerGlyTyrAlaLysThrLysAsnTyr 108
QY 198 GAATTGAACGATAAGCTGCTAGCTTCATTAAAGCCCGGT 236
Db 109 LysAlaLysHisGluLysSerTyrPheValSerProGly 121

RESULT 10
Q8SNH2 ID Q8SNH2 PRELIMINARY; PRT; 168 AA.
AC Q8SNH2;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE MHC class I alpha antigen (Fragment).
OS Xiphophorus hellerii guentheri.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percormorpha; Atherinomorpha;
OC Cyprinodontiformes; Poeciliidae; Xiphophorus.
OX NCBI_TaxID=188717;
RN [1]
RP SEQUENCE FROM N.A.
RA Fiqueroa F., Mayer W.E., Sato A., Zaleska-Rutczynska Z., Hess B.,
  Tichy H., Klein J.;
RT "MHC class I genes of swordtail fishes, Xiphophorus: variation in the
  number of loci and existence of ancient gene families.";
RL Immunogenetics 53:695-708(2001).
DR EMBL; AF411653; AAM18399.1; -.
FT NON_TER 1
FT NON_TER 168
SQ SEQUENCE 168 AA; 19734 MW; 5546EC2B83FB8009 CRC64;

Alignment Scores:
Pred. No.: 2.82 Length: 168
Score: 75.00 Matches: 16
Percent Similarity: 52.31% Conservative: 18
Best Local Similarity: 24.62% Mismatches: 29
Query Match: 12.40% Indels: 2
DB: 7 Gaps: 2

US-09-856-221-4 (1-332) x Q8SNH2 (1-168)

QY 54 CAGCAGCAGCTGGCTTCCTACAAAGCAAAATTCAGC---AATACAGCGTTGTATACTGG 110
Db 23 GluValGlnIleSerTyrCysAspSerArgThrAsnGluAsnIleProLysGlnAsnTrp 42
QY 111 CTACGTGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 170
Db 43 Met---AsnLysValSerSerGluTyrProAspTyrIleLysGluLysGluThrCys 61
QY 171 TTGATGCTGAACAAACTTACCAGTAGTAAATGAACGATAAGCTGTACCTTCATTAAG 230
Db 171 TTGATGCTGAACAAACTTACCAGTAGTAAATGAACGATAAGCTGTACCTTCATTAAG 230

```

```
Db 62 LeuValLysGlnThrPheLysTyrAsnLeuGluIleAlaLysThrArgPheAsnGln 81
QY 231 CCCGGTGGCTGGCAT 245
DB 82 ThrGlyGlyValHis 86
RESULT 11
Q9L815 PRELIMINARY; PRT; 598 AA.
AC Q9L815;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE TprD3.
GN TPRD.
OS Treponema pallidum (subsp. pertenue) (Yaws treponeme).
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=168;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CAUTHIER;
RX MEDLINE=20200377; PubMed=10735882;
RA Centurion-Lara A., Sun E.S., Barrett L.K., Castro C., Lukehart S.A.,
  Van Voorhis W.C.;
RT "Multiple alleles of treponema pallidum repeat gene b in treponema
  pallidum isolates."
RL J. Bacteriol. 182:2332-2335(2000).
DR EMBL: AF217538; AAF64130.1; -.
DR InterPro: IPR000847; HTH_LysR.
DR InterPro: IPR003872; WOSP_C.
DR InterPro: IPR003857; WOSP_Nterm.
DR Pfam: PF02722; WOSP_C; 1.
DR Pfam: PF02707; WOSP_N; 1.
DR PROSITE: PS00044; HTH_LYSR_FAMILY; UNKNOWN.1.
SQ SEQUENCE 598 AA; 64562 MW; 60D1BE78D0A32278 CRC64;

Alignment Scores:
Pred. No.: 3.46 Length: 598
Score: 74.50 Matches: 36
Percent Similarity: 38.13% Conservative: 17
Best Local Similarity: 25.90% Mismatches: 40
Query Match: 12.31% Indels: 46
DB: 2 Gaps: 5

US-09-856-221-4 (1-332) x Q9L815 (1-598)
QY 53 TCAGGCACAGCTGGCTTCCTACAAAGCAATTCAGCAATACAGCTGTGTAATACTGG-- 110
DB 19 SerGlyTyrAlaGlyValLeuThrProGlnVal-SerGlyThrAlaGlnLeuGlnTrpG1 38
QY 111 -----CTACGTGGCGCATTTGGCGGCTATTATTATCA 142
DB 38 yIleAlaPheGlnLysAsnProHisThrValProGlyLysHisThrHisGlyPheArgTh 58
QY 143 GYTTATGACTGGCTGTTCCCTGCTTTGATGCTGCTGCAAACTTAC----- 191
DB 58 rThrAenSerLeuThrIleSerLeuProLeuValSerLysHisThrHisThrArgArgG1 78
QY 192 -----CAGTATGAATTGAACGATAAGCTGTACGC----- 221
DB 78 yGluAlaArgSerGlyValTrpAlaGlnLeuGlnLeuLysAspLeuAlaValGluLeuAl 98
QY 222 -----TTCATTAGCCCGTGGC----- 239
DB 98 aSerSerLysSerThrAlaLeuSerPheThrLysProThrAlaSerPheGlnAlaTh 118
QY 240 -----TGGCATGGCACTTATGCTGTTGTTAGCAGGTGAACCTTGATGCTGAATTT 292
DB 118 rLeuHisCysTyrGlyAlaTyrLeuThrValGlyThrSerProSerCysValValAsnPh 138
QY 293 GGCACAGATGGAAAA-----AACTATTGGGAAGAGATGAACG 332
DB 138 eAlaGlnLeuTrpLysProPheValThrArgAlaTyrSerGluLysAspThrArg 156

RESULT 12
O58431 PRELIMINARY; PRT; 439 AA.
AC O58431;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical protein PH0700.
GN PH0700.
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococci; Thermococcaceae; Thermococcaceae;
OX Pyrococcus.
OX NCBI_TaxID=53953;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OT3;
RX MEDLINE=98344137; PubMed=9679194;
RA Kavarabavasi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
  Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
  Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
  Funahashi T., Tanaka T., Kudo Y., Yamazaki J., Kushida N., Oguchi A.,
  Aoki K.-I., Ioshizawa T., Nakamura Y., Robb F.I., Horikoshi K.,
  Masuchi Y., Shizuya H., Kikuchi H.;
RT "Complete sequence and gene organization of the genome of a hyper-
  thermophilic archaeobacterium, Pyrococcus horikoshii OT3."
RL DNA Res. 5:55-76(1998).
DR EMBL: AP000003; BAA29791.1; -.
KW Hypothetical protein; Complete proteome
SQ SEQUENCE 439 AA; 48361 MW; EB9F404553B15362 CRC64;

Alignment Scores:
Pred. No.: 5.17 Length: 439
Score: 73.00 Matches: 18
Percent Similarity: 55.56% Conservative: 12
Best Local Similarity: 33.33% Mismatches: 20
Query Match: 12.07% Indels: 4
DB: 17 Gaps: 2

US-09-856-221-4 (1-332) x O58431 (1-439)
QY 132 ATTATTATCATGTTTAT-----GACTGGCTGTTCCTGTGTGTGATGGGTGAA 182
DB 122 IlePheTyrMetPhePheLeuAlaAlaThrIleProLeuProLeuSerLeuIleSerArg 141
QY 183 CAACCTTACCAAGTATGAATGAACGATAAAGCTGTACGCTTCATTAAAGCCGCTGCTG 242
DB 142 GluPheArgArgTyrGluIleAsnLysAlaIleGlyArgPheAsnGluLeuGlyGlyTrp 161
QY 243 CATGCACTTATGCTGTTTGTAGCAGGTGAACCTTGATG 284
DB 162 ---GlyTrpValValGlyLeuIleLeuGlyPheThrLeuIle 174

RESULT 13
Q9UW87 PRELIMINARY; PRT; 1606 AA.
AC Q9UW87;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE MRP-like transporter.
GN MRP1.
OS Candida albicans (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OX NCBI_TaxID=5476;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. SS;
RX MEDLINE=21927371; PubMed=11929516;
RA Theiss S., Kretschmar M., Nichterlein T., Hof H., Agabian N.,
  Hacker J., Koehler G.A.;
RT "Functional analysis of a vacuolar ABC transporter in wild-type
```

QY 147 TATGACTTG-----GCTGTTTCCTGTGTTTGATGGCTGAACAAACTTACCAG--- 194

Alignment Scores:		
Pred. No.:	6.66	Length: 269
Score:	72.00	Matches: 24
Percent Similarity:	45.00%	Conservative: 32
Best Local Similarity:	30.00%	Mismatches: 32
Query Match:	11.90%	Indels: 12
DB:	16	Gaps: 3

```

Db 211 ArgAspAlaValAsnSerAlaSerSerLeuGlyLysSerSerLysSerSerTyrGlnMet 230
QY 195 -----TATGAATTGAACGATAAAGCT 215
Db 231 SerProLysAsnIleAsnGluAlaLeuHisGluValAlaMetAspIleAsnGluGlyAla 250
QY 216 ---GTACGCTTCATTAAACCGCGGCTGCCAT----- 245
Db 251 AspIleValMetValLysProGlyMetProTyrLeuAspIleIleSerLysValLysGlu 270
QY 246 -----GGCAGTTATGCTGGTTCTTAGCAGGTGAACCTTTCAGTCTCAATTTC 293
Db 271 ThrPheLysValProThrPheAlaTyrGlnValSerGlyGluTyrSerMetLeuLysLeu 290
QY 294 GCACAGATGGAAAAAAGCTTTGGAAAAAGAT 326
Db 291 Ala---IleAspLysGlyTrpLeuGluSerAsp 300

```

Search completed: January 15, 2003, 15:54:41  
Job time : 39.2773 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: January 15, 2003, 15:39:33 ; Search time 10.6207 Seconds  
(without alignments)  
3506.895 Million cell updates/sec

Title: US-09-856-221-1  
Perfect score: 793  
Sequence: 1 tgtcggcgcacacattat.....gtgacccaactgatacagtt 449

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 225784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame\_n2p.model -DEV=xlp  
-Q/cgn21/USFTO\_spool/US09856221/runat\_15012003\_153924\_12445/app\_query.fasta\_1.1948  
-DB=SwissProt\_40 -QFMT=fastn -SUFFIX=rs -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09856221 -CGN\_1\_1\_38 -runat\_15012003\_153924\_12445 -ICPU=3  
-NO\_XLPXY -NO\_WMAP -LARGESQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt\_40.\*

pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	78	9.8	468	ALS2_CANAL	O74657 candida alb
2	78	9.8	640	UL06_HSV7J	P52455 human herpe
3	76	9.6	318	TF2D_MESAU	P53360 mesocricetu
4	76	9.6	2555	PDS3_BACSU	P39847 bacillus su
5	76	9.6	2799	EDD_HUMAN	O95071 homo sapien
6	75	9.5	316	TF2D_MOUSE	P29037 mus musculu
7	75	9.5	1080	HDA4_CHICK	P83038 gallus gall
8	73.5	9.3	297	TF2D_XENLA	P27633 xenopus lae
9	73.5	9.3	339	TF2D_HUMAN	P20226 homo sapien
10	73.5	9.3	622	PKNE_MYCLE	P04575 mycobacteri
11	73.5	9.3	1718	RRPO_SHVX	O04575 shallot vir
12	73	9.2	256	YJJP_ECOLI	P39402 escherichia
13	72.5	9.1	805	E2F_DROME	O27368 drosophila
14	72.5	9.1	2688	ZEPI_MOUSE	O03172 mus musculu
15	72	9.1	442	ODO2_RAT	O01205 rattus norv
16	72	9.1	488	CWF8_SCHPO	O31011 schizosacch
17	72	9.1	881	GLND_VIBCH	Q9kpw0 vibrio chol
18	72	9.1	1145	PR22_YEAST	P24384 saccharomyc

19	71.5	9.0	201	1	RR4_SPIOL	P13788 spinacia ol	
20	71.5	9.0	470	1	RAN1_SCHPO	P08092 schizosacch	
21	71	9.0	300	1	TF2D_TRIFL	Q92117 trimeresuru	
22	71	9.0	302	1	TF2D_TRIGA	Q92146 trimeresuru	
23	71	9.0	509	1	YB44_CAEEL	Q10051 caenorhabdi	
24	71	9.0	886	1	YF1Q_ECOLI	P76594 escherichia	
25	71	9.0	2441	1	CBP_MOUSE	P45481 mus musculus	
C	26	70	481	1	MURC_TREPA	O83361 treponema p	
27	70	8.8	614	1	TUL1_SCHPO	Q09715 schizosacch	
28	70	8.8	624	1	PPSL_CAVPO	O54820 c bifunctio	
29	70	8.8	2303	1	POLG_TMEVG	P08545 t genome po	
30	69.5	8.8	368	1	ROX1_YEAST	P25042 saccharomyc	
31	69.5	8.8	373	1	CIS2_MYCTU	Q10529 mycobacteri	
C	32	69.5	601	1	NGPR_SALTU	P19618 salmo trutt	
33	69.5	8.8	2832	1	NDVB_RHIME	P20471 rhizobium m	
34	69	8.7	201	1	CWF1_HUMAN	Q16619 homo sapien	
35	69	8.7	302	1	TF2D_CHICK	Q13270 gallus gall	
36	69	8.7	381	1	ASG1_YEAST	P38986 saccharomyc	
37	69	8.7	423	1	MTB1_BACAM	P23941 bacillus am	
38	69	8.7	2303	1	POLG_TMEVB	P08544 t genome po	
C	39	68.5	8.8	396	1	ISPI_CAEEL	Q17848 caenorhabdi
40	68.5	8.6	1260	1	ALS1_CANAL	P46590 candida alb	
41	68.5	8.6	1419	1	ALAL_CANAL	Q13368 candida alb	
C	42	68.5	8.8	1659	1	VIT_ONCMY	O92093 oncorhynch
43	68	8.6	129	1	RL7_CHLMI	P38001 chlamydia m	
44	68	8.6	447	1	DNAI_SYNIF	P49995 synecocyst	
45	68	8.6	448	1	YAP1_CHICK	P46936 gallus gall	

ALIGNMENTS

RESULT 1	ALS2_CANAL	STANDARD;	PRT;	468 AA.
ID	ALS2_CANAL			
AC	O74657;			
DT	15-JUL-1999 (Rel. 38, Created)			
DT	15-JUL-1999 (Rel. 38, Last sequence update)			
DT	15-JUL-1999 (Rel. 38, Last annotation update)			
DE	Agglutinin-like protein 2 precursor (Fragment).			
GN	ALS2.			
OS	Candida albicans (Yeast).			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
OC	Saccharomycetales; mitosporic Saccharomycetales; Candida.			
OX	NCBI_TaxID=5476;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=1161;			
RX	MEDLINE=98440424; PubMed=9765564;			
RA	Hoyer L.L., Payne T.L., Hecht J.E.;			
RT	"Identification of Candida albicans ALS2 and ALS4 and localization of			
RT	als proteins to the fungal cell surface.";			
RL	J. Bacteriol. 180:5334-5343(1998).			
CC	-!- FUNCTION: MAY PLAY A ROLE IN ADHESION AND PATHOGENESIS.			
CC	-!- PFM: N-GLYCOSYLATED AND O-GLYCOSYLATED (POTENTIAL).			

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

EMBL; AF024580; AAC64235.1; - Repeat; Signal.  
Cell adhesion; Glycoprotein; POTENTIAL.  
FT SIGNAL 18 >468 AGGLUTININ-LIKE PROTEIN 2.  
FT CHAIN 1 253 253 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 315 315 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT NON\_TER 468 468  
SQ SEQUENCE 468 AA; 50203 MW; BFE773E169ED0FAF CRC64;

Alignment Scores:

```
Pred. No.: 2.75 Length: 468
Score: 78.00 Matches: 42
Percent Similarity: 42.18% Conservatives: 20
Best Local Similarity: 28.57% Mismatches: 53
Query Match: 9.84% Indels: 32
DB: 1 Gaps: 7

US-09-856-221-1 (1-449) x ALS2_CANAL (1-468)
QY 42 ATTGGATACGATTGGTGCAGAACCCGTTTGA---AGAAAGCGCGAGTCAACACCATCG 98
Db 298 ValGlySerSerLeuGlnSerLysProPheAsnLeuArgLeuArgGlyTyrAsnAsnSer 317
QY 99 TGATCATTTACGATAAAATGCGCAACTGCGTCAACGCGCGCGCTTGCAGCGGAAACG 158
Db 318 GluAlaAsnSerAsnGlyPheValIleValAlaThrThrArgThrValThrAspSerThr 337
QY 159 TACTGCAAAATTCGTTAACCGCTTGTTCCTCTCAGATAAACAACAACTGCAAGCTTA 218
Db 338 ThrAlaValThrThrLeuProPheAsnProSerValAspLysThrLysThrIleGluIle 357
QY 219 CTGGCAGAGCTTAGCACAAACCCATATAACTTACGTCAATAATCTGACAAATGTAGTCA 278
Db 358 LeuGlnProIleProThrIle-----ThrThr-----Ser 369
QY 279 GCGTGTGTCATTACCCATCTATGCGACACGACGATCGCGTACTGCTAGTGTGTCG 338
Db 370 TyrValGlyValThr-----ThrSerTyrSerThrLysThrAla----- 382
QY 339 GGTACCGCTCACAAGCGGAGGGATTGCTCGGACAGTAATGCGGATGACCGTTT 398
Db 383 -----ProIleGlyGluThrAlaThrValIleValAspValProTyr 396
QY 399 -----TCGGATTATTCGGA-----AAATGCCAAGTGGG 428
Db 397 HisThrThrThrThrValThrSerGluTrpThrGlyThrIleThrThrThrThrArg 416
QY 429 AGTGACCCCACTGATACATT 449
Db 417 ThrAsnProThrAspSerIle 423

RESULT 2
ID UL06_HSV7J STANDARD; PRT; 640 AA.
AC P52455;
DT 01-OCT-1996 (Rel. 34, Created)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Virion protein U76.
GN U76.
OS Human herpesvirus (type 7 / strain J1) (HHV7).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Simplexvirus.
OX NCBI_TaxID=57278;
RN [1]
RP SEQUENCE FROM N.A.
RA Nicholas J.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: PRESUMED VIRION PROTEIN; POSSIBLE ROLE IN DNA
CC PACKAGING.
CC -!- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL6,
CC EBV-1 56, EBV BRF1, HCMV UL104, AND VZV 54.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U43400; AAC54737.1;
CC InterPro; IPR002660; Herpes_UL6.
CC DR
```

```
DR Pfam: PF01763; Herpes_UL6; 1.
DR ProDom: PD003210; Herpes_UL6; 1.
SQ SEQUENCE 640 AA; 74538 MW; 4021A6E1B453FB63 CRC64;

Alignment Scores:
Pred. No.: 2.86 Length: 640
Score: 78.00 Matches: 25
Percent Similarity: 41.67% Conservatives: 15
Best Local Similarity: 26.04% Mismatches: 30
Query Match: 9.84% Indels: 26
DB: 1 Gaps: 3

US-09-856-221-1 (1-449) x UL06_HSV7J (1-640)
QY 37 TTTTCATTGGATACGATTGGTGCAGAACCCGTTTGAAGAACGCGCGCTTGCCTT 96
Db 441 PheThrLeuAspSerAspIleGluSerLeuSerLeuAlaLysAla----- 455
QY 97 CGTGATCATTTACGATATAAATGCGCAACTGCGTCAACGCGCGCTTGCCTT 150
Db 456 -----LeuAsnLysValGlnSerLeuProPheThr 465
QY 151 -----GCGAAACGTACTGCAAAATTCGTTAACCGCTTGTTCCTTCTT 192
Db 466 SerValSerIleAspAspThrArgAlaValAlaAsnSerPheSerGlnTyrIlePro 485
QY 193 CAGATA-----AACAAACAACTGCAAACTTACTGTCAGACGCTTAGCACAACGCTA 243
Db 486 AspileGluTyrAlaAspLysLysIleAspGlnLeuTrpGluThrGluTyrThrArgThr 505
QY 244 TATAACTTACGTCAATACTGACAAATGTATGATGTCAGCCGCTTGTTCATTA 291
Db 506 PheArgLeuArgLysAsnValAsnGlnGlyGlnGluAspSerIle 521

RESULT 3
ID TF2D_MESAU STANDARD; PRT; 318 AA.
AC P53360;
DT 01-OCT-1996 (Rel. 34, Created)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Transcription initiation factor TFIID (TATA-box factor) (TATA
DE sequence-binding protein) (TBP).
GN TBP OR TFIID.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN [1]
RP SEQUENCE FROM N.A.
RA Nouchi E.;
RL TISSUE=Kidney;
RT "The minimum essential region of CCG1/TA250 required for promoter-
RT selective transcription with TFIID complex.";
RL Thesis (1994), Kyushu University / Fukuoka, Japan.
CC -!- FUNCTION: GENERAL FACTOR THAT PLAYS A MAJOR ROLE IN THE ACTIVATION
CC OF EUKARYOTIC GENES TRANSCRIBED BY RNA POLYMERASE II. TFIID BINDS
CC SPECIFICALLY TO THE TATA BOX PROMOTER ELEMENT WHICH LIES CLOSE TO
CC THE POSITION OF TRANSCRIPTION INITIATION.
CC -!- SUBUNIT: BINDS DNA AS A MONOMER.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: BELONGS TO THE TBP FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D30051; BAA06287.1;
CC DR
```

DR HSP; P20226; LTGH.  
DR InterPro: IPR000814; TFIID.  
DR Pfam: PF00352; TBP; 2.  
DR PRINTS; PR00686; TFIIDR1D.  
DR PROSITE; PS00351; TFIID; 2.  
KW Transcription regulation; DNA-binding; Nuclear protein; Repeat.  
FT DOMAIN 59 72  
FT REPEAT 144 220 1.  
FT REPEAT 234 311 2.  
SQ SEQUENCE 318 AA; 34956 MW; A25FCD27A6422593 CRC64;  
Alignment Scores:  
Pred. No.: 4.28 Length: 318  
Score: 76.00 Matches: 26  
Percent Similarity: 43.27% Conservative: 19  
Best Local Similarity: 25.00% Mismatches: 45  
Query Match: 9.58% Indels: 14  
DB: 3 Gaps: 3  
US-09-856-221-1 (1-449) x TF2D\_MESAU (1-318)  
QY 70 TTAGAAGCGCGCGCTGATCAACCATTCGTGATCATATCATGATCAATGCGCACTG 129  
Db 52 LeuGluGlu-----GlnGlnArgGlnGlnGlnGlnGlnGlnGlnGlnGln 69  
QY 130 CGTCAACGCGCGCTGCGCGGAAAGCTACTGCAATTCGTTAACCGCTTGTTCCTT 189  
Db 70 GlnGlnGlnAlaValAlaThrAlaAlaLaserValGlnGlnSerThrSer----- 86  
QY 190 CCTCAGATAAACAACAACTGCAAGTACTGGCAGAGCTTACGACACAGCGCTATTAAC 249  
Db 87 -----GlnGlnSerThrGlnGlnAlaSerGlnGlnThrProGln 99  
QY 250 TTACGTCATATCTGACAAATGATGTCAGCGCTG-----TCATTACCATCATCG 303  
Db 100 LeuPheHisSerGlnThrLeuThrAlaProLeuProGlnThrProLeuThrPro 119  
QY 304 ACACGACGAGATCGCTCGCTAGCTGCTAGTGTGCGTCACGCGCTCAACGCGAGG 363  
Db 120 SerProMetThrProMetThrProIleThrProAlaThrProAlaSerGluSerGly 139  
QY 364 GATTTGCGCTCGG 375  
Db 140 IleValProGln 143  
RESULT 4  
ID\_PPS3\_BACSU STANDARD; PRT: 2555 AA.  
AC P39847;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Peptide synthetase 3.  
GN PPS3 OR PPS3.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=1423;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RA MEDLINE=98044033; PubMed=9384377;  
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,  
RA Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,  
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,  
RA Denisot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,  
RA Entlan K.D., Erlington J., Fabret C., Ferrari E., Foulger D.,  
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
RA Gim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,  
RA Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,  
RA Hilbert H., Hoisappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,

RA Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,  
RA Kurita K., Lapinus A., Lardinois S., Lauber J., Lazarevic V.,  
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,  
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,  
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
RA Rieger M., Rivolta C., Rocha E., Rochet B., Rose M., Sadade Y.,  
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
RA Sekiguchi J., Sekowska A., Seror S.J., Serrif P., Shin B.S., Soldo B.,  
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,  
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,  
RA Viari A., Wambutt R., Wedler E., Wedler K., Weitzenecker T.,  
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,  
RT "The complete genome sequence of the Gram-positive bacterium Bacillus  
RT subtilis";  
RL Nature 390:249-256(1997).  
RN [2]  
RP SEQUENCE OF 1-859 FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=95227362; PubMed=7711903;  
RA Tognoni A., Franchi E., Magistrelli C., Colombo E., Cosmina P.,  
RA Grandi G.;  
RT "A putative new peptide synthase operon in Bacillus subtilis: partial  
RT characterization";  
RL Microbiology 141:645-648(1995).  
CC -1- COFACTOR: CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTHETINES (BY  
CC SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME  
CC FAMILY.  
CC -1- SIMILARITY: CONTAINS 2 ACYL CARRIER DOMAINS.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: Z99113; CAB13715.1; -;  
CC DR EMBL: Z34883; CAB4362.1; -;  
CC DR HSP: P14687; LAMU.  
CC DR Subtilisin; BG10972; ppsC.  
CC DR InterPro: IPR000873; AMP-bind.  
CC DR InterPro: IPR001242; Condensatn.  
CC DR InterPro: IPR003880; Ppantne\_attach.  
CC DR Pfam: PF00501; AMP-binding; 2.  
CC DR Pfam: PF00550; pp-binding; 2.  
CC DR Pfam: PF00668; Condensation; 3.  
CC DR PRINTS; PR00154; AMPBINDING.  
CC DR PROSITE; PS00012; PHOSPHOPANTHETINE; 1.  
CC DR PROSITE; PS00455; AMP BINDING; 2.  
CC DR PROSITE; PS00075; ACP DOMAIN; 2.  
CC KW Multifunctional enzyme; Ligase; Repeat; Complete proteome.  
CC FT DOMAIN 972 1039 ACYL CARRIER (ACP) 1.  
CC FT DOMAIN 2008 2074 ACYL CARRIER (ACP) 2.  
CC FT BINDING 2038 2038 PHOSPHOPANTHETINE (POTENTIAL).  
CC SQ SEQUENCE 2555 AA; 287499 MW; 3E50B3395105D5D0 CRC64;  
Alignment Scores:  
Pred. No.: 5.48 Length: 2555  
Score: 76.00 Matches: 43  
Percent Similarity: 34.86% Conservative: 18  
Best Local Similarity: 24.57% Mismatches: 44  
Query Match: 9.58% Indels: 70  
DB: 1 Gaps: 8  
US-09-856-221-1 (1-449) x PPS3\_BACSU (1-2555)  
QY 52 GATGTGTCAGAACCCCGTTTAGAAGAGCGCGAGTCACCAACCATTCGT----- 99

Db 2234 AspleuSerAlaGlyTyrGlnGlnAlaAlaAlaGlyGlnThrGlnLeuProProLys 2253  
 Qy 100 ---GATCATATTATGACGATGCGGCAACTGCGTCAACCGCGGCTTGCGGGGAAA 156  
 Db 2254 ThrAspSerTyrGlnGluTyrAlaArgAlaGlnGluTyrAlaA 2268  
 Qy 157 CGTACGCAAAATTCGTTAAACCGCTTCTCCTCCAGATAAACAACAACTG 210  
 Db 2269 ---GATCATATTATGACGATGCGGCAACTGCGTCAACCGCGGCTTGCGGGGAAA 2276  
 Qy 211 CAAAGTACTGCGT---GATCATATTATGACGATGCGGCAACTGCGTCAACCGCGGCTTGCGGGGAAA 2282  
 Db 2277 GluAlaTyrPargSerValGluGlnGlnAlaAlaGluLeuProTyrGluIlePro 2296  
 Qy 256 CATATCTGACAAATGAT---GGTCAGCCGCTGTCATACCCATCTATGCGACACAGCA 312  
 Db 2297 HisHisValAsnIleAspPheSerLysArgAspSerLysPheSerLeuThrGluAla 2316  
 Qy 313 GATCCCTCGTACTGCTT---GATCATATTATGACGATGCGGCAACTGCGTCAACCGCGGCTTGCGGGGAAA 330  
 Db 2317 AspThrAlaValLeuLeuGlnAsnValAsnHisAlaTyrGlyThrAspThrGlnAspIle 2336  
 Qy 331 ---GATCATATTATGACGATGCGGCAACTGCGTCAACCGCGGCTTGCGGGGAAA 363  
 Db 2337 LeuLeuThrAlaAlaSerLeuAlaIleCysGluTyrThrGlyGlySerLysLeuArgIle 2356  
 Qy 364 ---GATCATATTATGACGATGCGGCAACTGCGTCAACCGCGGCTTGCGGGGAAA 381  
 Db 2357 AlaMetGluGlyHisGlyArgGluHisIleLeuProGluLeuAspIleSerArgThrVal 2376  
 Qy 382 ---ATGCCGATGACGCTTTTCCGATATTCCTGGAAT 417  
 Db 2377 GlyTyrPheThrSerMetTyrProAlaLeuLeuSerPheGluAsn 2391  
 RESULT 5  
 EDD\_HUMAN  
 ID EDD\_HUMAN STANDARD; PRT: 2799 AA.  
 AC Q950J1; Q9NPL3; Q9A970;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Ubiquitin-protein ligase EDD (EC 6.3.2.-) (Hyperplastic discs  
 protein homolog) (hHYD) (Progestin induced protein).  
 GN EDD OR HYD OR KIAA0896.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Placenta, and Heart;  
 RX MEDLINE=99153743; PubMed=10030672;  
 RA Callaghan M.J., Russell A.J., Woollett E., Sutherland G.R.,  
 RA Sutherland R.L., Watts C.K.W.;  
 RT "Identification of a human HECT family protein with homology to the  
 RT Drosophila tumor suppressor gene hyperplastic discs";  
 RL Oncogene 17:3479-3491(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Fetal brain;  
 RX MEDLINE=21671350; PubMed=11714696;  
 RA Honda Y., Tojo M., Matsuzaki K., Anan T., Matsumoto M., Ando M.,  
 RA Saya H., Nakao M.;  
 RT "Cooperation of HECT-domain ubiquitin ligase hHYD and DNA  
 RT topoisomerase II-binding protein for DNA damage response";  
 RL J. Biol. Chem. 277:3599-3605(2002).  
 RN [3]  
 RP SEQUENCE OF 1569-2799 FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=99156230; PubMed=10048485;  
 RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirose M.,  
 RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;

RT "Prediction of the coding sequences of unidentified human genes. XII.  
 RT The complete sequences of 100 new cDNA clones from brain which code  
 RT for large proteins in vitro.";  
 RL DNA Res. 5:355-364(1998).  
 RN [4]  
 RP X-RAY CRYSTALLOGRAPHY (1.04 ANGSTROMS) OF 2391-2455, AND MUTAGENESIS  
 RP OF CYS-2768.  
 RX MEDLINE=21192643; PubMed=11287654;  
 RA Deo R.C., Sonenberg N., Burley S.K.;  
 RT "X-ray structure of the human hyperplastic discs protein: an ortholog  
 RT of the C-terminal domain of poly(A)-binding protein.";  
 RL Proc Natl Acad Sci U S A 98:4414-4419(2001)  
 CC -1- FUNCTION: E3 ubiquitin-protein ligase which accepts ubiquitin from  
 CC an E2 ubiquitin-conjugating enzyme in the form of a thioester and  
 CC then directly transfers the ubiquitin to targeted substrates (By  
 CC similarity). This protein may be involved in maturation and/or  
 CC post-transcriptional regulation of mRNA. May play a role in  
 CC control of cell cycle progression. May have tumor suppressor  
 CC function. Regulates DNA topoisomerase II binding protein (TopBP1)  
 CC for the DNA damage response.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- TISSUE SPECIFICITY: Widely expressed. Most abundant in testis and  
 CC expressed at high levels in brain, pituitary and kidney.  
 CC -1- MISCELLANEOUS: A cysteine residue is required for ubiquitin-  
 CC thioester formation.  
 CC -1- SIMILARITY: CONTAINS 1 HECT-TYPE E3 UBIQUITIN-PROTEIN LIGASE  
 CC DOMAIN.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC  
 CC EMBL: AF006010; AAD01259.2; --  
 CC EMBL: U95000; AAF88143.1; --  
 CC DR EMBL; AB020703; BAA74919.1; --  
 CC PDB: 1I2T; 18-APR-02.  
 CC DR InterPro: IPR000569; HECT\_domain.  
 CC DR InterPro: IPR002004; PABP/HECT.  
 CC DR InterPro: IPR000449; UBA\_domain.  
 CC DR InterPro: IPR003126; Znf\_Nreoguin.  
 CC DR Pfam: PF00627; UBA; 1.  
 CC DR Pfam: PF00632; HECT; 1.  
 CC DR Pfam: PF00658; PABP; 1.  
 CC DR Pfam: PF02207; zf-UBRL; 1.  
 CC DR SMART; SM00119; HECTC; 1.  
 CC DR SMART; SM00517; POLYA; 1.  
 CC DR SMART; SM00396; Znf\_UBRL; 1.  
 CC DR PROSITE; PS00237; HECT; 1.  
 CC DR Ubl conjugation pathway; Ligase; Nuclear protein; 3D-structure.  
 CC KW DOMAIN 2393 2449 PABP-LIKE.  
 CC FT DOMAIN 2393 2449 HECT.  
 CC FT DOMAIN 2462 2799 ASP/GLU-RICH (ACIDIC).  
 CC FT DOMAIN 1986 1997 PRO-RICH.  
 CC FT DOMAIN 2036 2059 ARG/ASP-RICH (MIXED CHARGE).  
 CC FT DOMAIN 2329 2348 ARG/ASP-RICH (MIXED CHARGE).  
 CC FT DOMAIN 2357 2366 ARG/ASP-RICH (MIXED CHARGE).  
 CC FT DOMAIN 2489 2500 ASP/GLU-RICH (ACIDIC).  
 CC FT DOMAIN 2737 2757 PRO-RICH.  
 CC FT DOMAIN 940 945 POLY-GLU.  
 CC FT DOMAIN 980 985 POLY-SER.  
 CC FT DOMAIN 1528 1537 POLY-SER.  
 CC FT DOMAIN 1671 1681 POLY-SER.  
 CC FT DOMAIN 1762 1768 POLY-ALA.  
 CC FT BINDING 2768 2768 UBIQUITIN (BY SIMILARITY).  
 CC FT MUTAGEN 2768 2768 C->A: LOSS OF UBIQUITIN BINDING.  
 CC FT CONFLICT 134 134 S -> P (IN REF. 2).  
 CC FT CONFLICT 229 229 E -> K (IN REF. 2).  
 CC FT CONFLICT 258 258 S -> Y (IN REF. 2).  
 CC FT CONFLICT 374 375 IG -> M (IN REF. 2).  
 CC FT CONFLICT 772 772 D -> H (IN REF. 2).

```

FT CONFLICT 780 780 Q -> R (IN REF. 2).
FT CONFLICT 884 884 D -> G (IN REF. 2).
FT CONFLICT 1811 1811 S -> P (IN REF. 2).
FT CONFLICT 2144 2144 L -> H (IN REF. 2).
FT CONFLICT 2282 2282 K -> R (IN REF. 2).
FT CONFLICT 2474 2474 MISSING (IN REF. 3).
FT CONFLICT 2489 2489 D -> N (IN REF. 2).
SQ SEQUENCE 2799 AA: 309348 MW: 871300DB404FF561 CRC64:

Alignment Scores:
Pred. No.: 5.54 Length: 2799
Score: 76.00 Matches: 33
Percent Similarity: 43.59% Conservative: 18
Best Local Similarity: 28.21% Mismatches: 46
Query Match: 9.58% Indels: 20
DB: 1 Gaps: 5

US-09-856-221-1 (1-449) x EDD_HUMAN (1-2799)

QY 43 TTGGATAACGATTGGTCAGAACCCGTTTAGAAGAGCCGCAAGTCAACACCATCGTGAT 102
Db 1409 LeuGlnAsnLysTyrThrProGlyArgGluGluAlaValThr----- 1425
QY 103 CATTATCAGCATAAATGCGGCAACTGGCTCAACGCGC-----GCCTTGCGG 150
Db 1426 -----MetArgPheLeuArg9SerValAlaArgValPheValIleLeuSer 1440
QY 151 GCGAAACGCTACTGCAATCGTTTAAACCGCTTTGTTCTCTCAGATAAACAAAACATG 210
Db 1441 ValGluMetAlaSerSerLysLysLysAsnAsnPhelProGlnProIleGlyLysCys 1460
QY 211 CAAAGTTACTGGCAGAGCTTA-----GCACAACGCTATATAACTTACGTGCAT 258
Db 1461 LysArgValPheGlnAlaLeuLeuProTyrAlaValGluGluLeuCysAsnValAlaGlu 1480
QY 259 AATCTGACATATGATGTGACGCGTGTCTATCCCATCTATGGACACACGACATCGG 318
Db 1481 SerLeuIleVal-----ProValArgMetGlyIle---AlaArgProThrAlaPro 1496
QY 319 TCCGTACGTCAGTGTGTCGCGTCACGCGCTCAACAGCGCGAGGGATTG 369
Db 1497 PheThrLeuAlaSerThrSerIleAspAlaMetGlnGlySerGluGluLeu 1513

RESULT 6
TF2D_MOUSE STANDARD; PRT; 316 AA.
AC P29037: 009169;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Transcription initiation factor TFIIID (TATA-box factor) (TATA
DE sequence-binding protein) (TBP).
GN TBP OR TFIIID.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID:10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-91319543; PubMed-1861978;
RA Tamura T.-A., Sumita K., Fujino I., Aoyama A., Horikoshi M.,
RA Hoffmann A., Roeder R.G., Muramatsu M., Mikoshiba K.;
RT *Striking homology of the 'variable' N-terminal as well as the
RT 'conserved core' domains of the mouse and human TATA-factors
RT (TFIIID).
RL Nucleic Acids Res. 19:3861-3865(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE-93324347; PubMed-8332475;
RA Sumita K., Makino Y., Katoh K., Kashimoto T., Muramatsu M.,
RA Mikoshiba K., Tamura T.-A.;
RT *Structure of a mammalian TBP (TATA-binding protein) gene: isolation
RT of the mouse TBP genome.

```

```

RL Nucleic Acids Res. 21:2769-2769(1993).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-TW18/TW5; TISSUE-Testis;
RA MEDLINE-97432814; PubMed-9286694;
RA Trachtulec Z., Hamvas R.M., Forejt J., Lehrach H.R., Vincek V.,
RA Klein J.;
RT *Linkage of TATA-binding protein and proteasome subunit C5 genes in
RT mice and humans reveals syntenic conserved between mammals and
RT invertebrates.
RL Genomics 44:1-7(1997).
CC -1- FUNCTION: GENERAL FACTOR THAT PLAYS A MAJOR ROLE IN THE ACTIVATION
CC OF EUKARYOTIC GENES TRANSCRIBED BY RNA POLYMERASE II. TFIIID BINDS
CC SPECIFICALLY TO THE TATA BOX PROMOTER ELEMENT WHICH LIES CLOSE TO
CC THE POSITION OF TRANSCRIPTION INITIATION.
CC -1- SUBUNIT: BINDS DNA AS A MONOMER.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: BELONGS TO THE TBP FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D01034; BAA00840.1; -
DR EMBL; U63933; AAB53097.1; -
DR PIR; S34437; S34437.
DR HSP; P20226; ITGH.
DR TRANSFAC; T00796; -.
DR MGD; MGI:101838; Tbp.
DR InterPro; IPR000814; TFIIID.
DR Pfam; PF00352; TBP; 2.
DR PRINTS; PR00686; TIFACTORIID.
DR PROSITE; PS00351; TFIIID; 2.
KW Transcription regulation; DNA-binding; Nuclear protein; Repeat.
FT DOMAIN 55 70 POLY-GLN.
FT REPEAT 142 218 1.
FT REPEAT 232 309 2.
FT VARIANT 70 70 Q -> QQ (IN STRAIN TW18/TW5).
SQ SEQUENCE 316 AA: 34709 MW: A33D776160B4A393 CRC64:

Alignment Scores:
Pred. No.: 5.46 Length: 316
Score: 75.00 Matches: 23
Percent Similarity: 43.16% Conservative: 18
Best Local Similarity: 24.21% Mismatches: 42
Query Match: 9.46% Indels: 12
DB: 1 Gaps: 2

US-09-856-221-1 (1-449) x TF2D_MOUSE (1-316)
QY 97 CGTGATCATATATCAGCATAAATCGGCAACTCGTCAACGGCGCGCTTCGCGCGAAA 156
Db 57 ArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 76
QY 157 CGTACTGCAGAAATTCGTTAAACGCGCTTGTCTCTCTCAGATAAACAACAACTGCAAGT 216
Db 77 AlaSerValGlnGlnSerThrSer-----GlnGln 86
QY 217 TACTGGCAGACGTTAGCAACAGCCCTATATACTTACGTCAATAATCTGCAATTTGATGT 276
Db 87 ProThrGlnGlnAlaSerGlyGlnThrProGlnLeuPheHisSerGlnThrLeuThr 106
QY 277 CAGCGGTG-----TCATTACCATCTATGCGACACACGACAGATCCGTCGCTACTCCT 330
Db 107 AlaProLeuProGlyThrThrProLeuTyrProSerProMetThrProMetThrProIle 126
QY 331 AGTGCTGCGCTCACCGCTCAACAGCGGAGGATTTGCCTCGG 375
Db 127 ThrProAlaThrProAlaSerGlnSerGlyIleValProGln 141

```

```
RESULT 7
ID HD4A_CHICK STANDARD; PRT: 1080 AA.
AC P83038;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Histone deacetylase 4 (HD4).
GN HDAC4.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RA Takechi S., Azuma R., Nakayama T.;
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: RESPONSIBLE FOR THE DEACETYLATION OF LYSINE RESIDUES ON
CC THE N-TERMINAL PART OF THE CORE HISTONES (H2A, H2B, H3 AND H4).
CC HISTONE DEACETYLATION PLAYS AN IMPORTANT ROLE IN TRANSCRIPTIONAL
CC REGULATION, CELL CYCLE PROGRESSION AND DEVELOPMENTAL EVENTS (BY
CC SIMILARITY). REPRESENTS TRANSCRIPTION (BY similarity).
CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -1- SIMILARITY: BELONGS TO THE HISTONE DEACETYLASE / ACUC / APHA
CC FAMILY. HD SUBFAMILY 2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AB052839; BAB60957.1;
DR InterPro: IPR000286; Hist.deacetylase.
DR Pfam: PF00850; Hist.deacetylase.
DR PRINTS: PR01270; HDASUPER.
KW Hydrolase; Nuclear protein; Transcription regulation; Repressor.
FT DOMAIN 651 1080 HISTONE DEACETYLASE.
SQ SEQUENCE 1080 AA; 119467 MW; B6416E2C43F1428C CRC64;

Alignment Scores:
Pred. No.: 6.31 Length: 1080
Score: 75.00 Matches: 28
Percent Similarity: 37.04% Conservative: 12
Best Local Similarity: 25.93% Mismatches: 35
Query Match: 9.46% Indels: 33
DB: 1 Gaps: 3

US-09-856-221-1 (1-449) x HD4A_CHICK (1-1080)
QY 78 AGCGCGCGTCAAAACATTCGTGATCATATCAGCATAAATGCGCAACTGCGTCAACG 137
Db 300 SerSerProAsnSerSerAsnAsnSerAlaGluAsnGlyIleThrGlySerVal 319
QY 138 CGCGCGCTTCCGCGGCAAGACTGCTGCAATTCGTTACCGCTTGTCTCTCTCAGAT 197
Db 320 ThrSerIleGlnAlaGluThr-Ser----- 327
QY 198 AAACAAAAAAGTCAAAAGTTACTGGCAGAGCTTAGCACAGCGCTATATACCTACGTCA 257
Db 328 -----LeuAlaHisArgLeuValAsnArg----- 335
QY 258 TAATCTGCAATGTATGTCAGCGTGTGATCATACCATCTATCGGCACACCATCC 317
Db 336 -----GluGlySerValThrGlnLeuProLeuTyThrSerProSerLeuPr 351
QY 318 GTCGCTACTGCTAGTGTGCGGTCACCGCTCACAAGCGCGAGGG-----GA 365
Db 351 oAsnIleThrLeuGlyLeuProAlaThrGlyProSerSerGlyGlySerAlaGlnGlnAs 371
```

```
QY 366 TTTCCTCGGACAGTAATCCCG 387
Db 371 PALAGLUArgLeuAlaIlePro 378

RESULT 8
ID TF2D_XENLA STANDARD; PRT: 297 AA.
AC P27633;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Transcription initiation factor TFIIID (TATA-box factor) (TATA
DE sequence-binding protein) (TBP).
GN TBP OR TFIIID.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RA Hashimoto S., Fujita H., Hasegawa S., Roeder R.G., Horikoshi M.;
RT "Conserved structural motifs within the N-terminal domain of TFIIID
RT tau from Xenopus, mouse and human."
RL Nucleic Acids Res. 20:3788-3788(1992).
CC -1- FUNCTION: GENERAL FACTOR THAT PLAYS A MAJOR ROLE IN THE ACTIVATION
CC OF EUKARYOTIC GENES TRANSCRIBED BY RNA POLYMERASE II. TFIIID BINDS
CC SPECIFICALLY TO THE TATA BOX PROMOTER ELEMENT WHICH LIES CLOSE TO
CC THE POSITION OF TRANSCRIPTION INITIATION.
CC -1- SUBUNIT: BINDS DNA AS A MONOMER.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: BELONGS TO THE TBP FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X66033; CAA46832.1;
DR HSSP: P20226; ITGH.
DR TRANSFAC: T02111.
DR InterPro: IPR000814; TFIIID.
DR Pfam: PF00352; TBP; 2.
DR PRINTS: PR00686; TIFACTORIID.
DR PROSITE: PS00351; TFIIID; 2.
KW Transcription regulation; DNA-binding; Nuclear protein; Repeat.
FT DOMAIN 54 63 POLY-GLN.
FT REPEAT 123 199 1.
FT REPEAT 213 290 2.
SQ SEQUENCE 297 AA; 32703 MW; 9893593418AEFD81 CRC64;

Alignment Scores:
Pred. No.: 7.8 Length: 297
Score: 73.50 Matches: 25
Percent Similarity: 48.15% Conservative: 14
Best Local Similarity: 30.86% Mismatches: 33
Query Match: 9.27% Indels: 9
DB: 1 Gaps: 3

US-09-856-221-1 (1-449) x TF2D_XENLA (1-297)
QY 157 CGTACTGCAATTCGTACCGCTTGTCTCTCTCAGATAACAAAACTGCAACT 216
Db 43 GlnThrThrAsnSerSerIleLeu---GluGlnGlnArgGlnGlnGlnThr 61
QY 217 TACTGGCAGACGTAGCACAA-----CGCCTATATACCTAGCTCAT 258
Db 62 GlnGlnSerThrLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnPheHis 81
```

QY 259 AATCTGACATTCATGCTCAGCGCTG-----TCATTACCATCATTCGCGACACCA 312  
 Db 82 ProGlnThrLeuThrAlaProLeuProGlyAsnThrProLeuThrProSerProTle 101  
 QY 313 GATCCGTCGCTACTGCTTAGTGGCTGCGCTCAGCGCTCACAAGCGGAGGGATTGCT 372  
 Db 102 ThrProMetThrProIleSerProAlaThrProAlaSerGluSerSerGlyIleValPro 121  
 QY 373 CGG 375  
 Db 122 Gln 122  
 RESULT 9  
 TF2D\_HUMAN STANDARD; PRT; 339 AA.  
 AC P20226; Q16845; Q9UC02;  
 DT 01-FEB-1991 (Rel. 17, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Transcription initiation factor TFIID (TATA-box factor) (TATA  
 DE sequence-binding protein) (TBP).  
 GN TBP OR TFIID OR TF2D.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Fibroblast;  
 RX MEDLINE=90302010; PubMed=2194289;  
 RA Kao C.C., Lieberman P.M., Schmidt M.C., Zhou Q., Pei R., Berk A.J.;  
 RT "Cloning of a transcriptionally active human TATA binding factor.";  
 RL Science 248:1646-1649(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90326195; PubMed=2374612;  
 RA Hoffmann A., Sinn E., Yamamoto T., Wang J., Roy A., Horikoshi M.,  
 RA Roeder R.G.;  
 RT "Highly conserved core domain and unique N terminus with presumptive  
 RT regulatory motifs in a human TATA factor (TFIID).";  
 RL Nature 346:387-390(1990).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Smalley C.;  
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP DOMAINS  
 RX MEDLINE=90302006; PubMed=2363050;  
 RA Peterson M.G., Tanese N., Pugh B.F., Tjian R.;  
 RT "Functional domains and upstream activation properties of cloned  
 RT human TATA binding protein.";  
 RL Science 248:1625-1630(1990).  
 RN [5]  
 RX X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 159-337.  
 RP MEDLINE=96209823; PubMed=8643494;  
 RA Nikolov D.B., Chen H., Halay E.D., Hoffmann A., Roeder R.G.,  
 RA Burley S.K.;  
 RT "Crystal structure of a human TATA box-binding protein/TATA element  
 RT complex.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 93:4862-4867(1996).  
 RN [6]  
 RX X-RAY CRYSTALLOGRAPHY (2.9 ANGSTROMS) OF 159-339.  
 RP MEDLINE=96346176; PubMed=8757291;  
 RA Joo T.S., Chiu T.K., Leiberman P.M., Balkalov I., Berk A.J.,  
 RA Dickerson R.E.;  
 RT "How proteins recognize the TATA box.";  
 RL J. Mol. Biol. 261:239-254(1996).  
 CC -1- FUNCTION: GENERAL FACTOR THAT PLAYS A MAJOR ROLE IN THE ACTIVATION  
 CC OF EUKARYOTIC GENES TRANSCRIBED BY RNA POLYMERASE II. TFIID BINDS  
 CC SPECIFICALLY TO THE TATA BOX PROMOTER ELEMENT WHICH LIES CLOSE TO  
 CC THE POSITION OF TRANSCRIPTION INITIATION.  
 CC -1- SUBUNIT: BINDS DNA AS A MONOMER.

CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- SIMILARITY: BELONGS TO THE TBP FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M55654; AAA36731.1; -  
 DR EMBL; M34960; AAC03409.1; -  
 DR EMBL; X54993; AAC38736.1; -  
 DR EMBL; AL031259; CAA20286.1; -  
 DR PIR; A34830; A34830.  
 DR PIR; A34831; A34831.  
 DR PDB; 1TGH; 01-AUG-96.  
 DR PDB; 1CDW; 23-DEC-96.  
 DR TRANSFAC; T00794; -  
 DR Genew; HGNC:11588; TBP.  
 DR MIM; 600075; -  
 DR InterPro; IPR000814; TFIID.  
 DR Pfam; PF00352; TBP; 2.  
 DR PRINTS; PR00686; TIFACTORIID.  
 DR PROSITE; PS00351; TFIID; 2.  
 KW Transcription regulation; DNA-binding; Nuclear protein; Repeat;  
 KW 3D-structure.  
 FT DOMAIN 55 95 POLY-GLN.  
 FT REPEAT 165 241 1.  
 FT REPEAT 255 332 2.  
 FT CONFLICT 58 61 MISSING (IN REF. 2).  
 FT CONFLICT 187 187 A -> R (IN REF. 1; AAC03409).  
 SQ SEQUENCE 339 AA; 37698 MW; A61A578D972B970B CRC64;  
 Alignment Scores:  
 Pred. No.: 7.93 Length: 339  
 Score: 73.50 Matches: 23  
 Percent Similarity: 42.72% Conservative: 21  
 Best Local Similarity: 22.33% Mismatches: 46  
 Query Match: 9.27% Indels: 13  
 Db: 1 Gaps: 2  
 US-09-856-221-1 (1-449) x TF2D\_HUMAN (1-339)  
 QY 73 GAAGAAGCGCCAGTCAAAACCATTCCTGATCATATCAGCATATAAATGCGCAACTGCGT 132  
 Db 73 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 92  
 QY 133 CAACGCGCGCGCTTGGCGGCAACGCTACTGCAAAATTCGTTAAGCGGCTTGTTCCTCCT 192  
 Db 93 GlnGlnGlnAlaValAlaAlaAlaAlaValGlnGlnSerThrSer----- 107  
 QY 193 CAGATAAACAACAAACTGCAAAAGTTACTGGGAGACGTTAGCAACACGCCCTATATACTTA 252  
 Db 108 -----GlnGlnAlaThrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 121  
 QY 253 CGTCATATTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 306  
 Db 122 PheHisSerGlnThrLeuThrThrAlaProLeuProGlyThrThrProLeuThrProSer 141  
 QY 307 CCACGATCGCTCGCTACTGCTTAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 366  
 Db 142 ProMetThrProMetThrProIleThrProAlaThrProAlaSerGluSerSerGlyIle 161  
 QY 367 TTGCTCGG 375  
 Db 162 ValProGln 164  
 RESULT 10  
 PKNB\_MYCLE  
 ID PKNB\_MYCLE STANDARD; PRT; 622 AA.  
 AC P54744;

DT 01-OCT-1996 (Rel. 34, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Probable serine/threonine-protein kinase pKns (EC 2.7.1.1.-).  
 GN PKNB OR ML0016.  
 OS Mycobacterium leprae.  
 OC Bacteria; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1769;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97124199; PubMed=8969512;  
 RA Fsihi H., de Rossi E., Salazar L., Cantoni R., Labo M.,  
 RA Riccardi G., Takiff H.E., Eiglmeyer K., Bergh S., Cole S.T.;  
 RT "Gene arrangement and organization in a approximately 76 kb fragment  
 RT encompassing the oric region of the chromosome of Mycobacterium  
 RT leprae.";  
 RL Microbiology 142:3147-3161(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=TN;  
 RX MEDLINE=21128732; PubMed=11234002;  
 RA Cole S.T., Eiglmeyer K., Parkhill J., James K.D., Thomson N.R.,  
 RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,  
 RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies K.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,  
 RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.M.,  
 RA Murphy L., Oliver L., Quail M.A., Rajandream M.A., Rutherford K.M.,  
 RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,  
 RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,  
 RA Barrell B.G.;  
 RT "Massive gene decay in the leprosy bacillus.";  
 RL Nature 409:1007-1011(2001).  
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; 270722; CAA94718.1; -;  
 DR EMBL; AL583917; CAC29524.1; -;  
 DR Leproma; ML0016; -;  
 DR InterPro; IPR000719; Euk\_pkinase.  
 DR InterPro; IPR002290; Ser\_thr\_pkinase.  
 DR InterPro; IPR001245; Tyr\_pkinase.  
 DR Pfam; PF00069; pkinase; 1.  
 DR ProDom; PD000001; Euk\_pkinase; 1.  
 DR SMART; SM00219; Tyrc; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
 DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 KW Transferase; Serine/threonine-protein kinase; ATP-binding;  
 KW Complete proteome.  
 FT DOMAIN 11 273 PROTEIN KINASE.  
 FT NP\_BIND 17 23 ATP (BY SIMILARITY).  
 FT BINDING 40 40 ATP (BY SIMILARITY).  
 FT ACT\_SITE 138 138 BY SIMILARITY.  
 FT CONFLICT 303 304 PG -> LC (IN REF. 1).  
 SQ SEQUENCE 622 AA; 66187 MW; 9611B0FD93F2BA7C CRC64;  
 Alignment Scores:  
 Pred. No.: 8.52 Length: 622  
 Score: 73.50 Matches: 30  
 Percent Similarity: 46.88% Conservative: 15  
 Best Local Similarity: 31.25% Mismatches: 34  
 Query Match: 9.27% Indels: 17  
 DB: 1 Gaps: 5  
 US-09-856-221-1 (1-449) x PKNB\_MYCLE (1-622)

QY 106 TATCAGCATAAATGCGCAACTGCGTCAACGCGCGGCTTGCAGGAAACG- 159  
 DB 225 TyrGlnHIs-----ValaTgGluAspProIleProProSerGlnaRgHIsGlu 240  
 QY 160 ---ACTGCAAAATTCGTTAAACCGCTTTGTTCCCTTCCTCACAATAAACAACCTCAAAAGT 216  
 DB 241 GlyIleSerValAspLeuAspAlaValValLeuLysAlaLeuAlaLysAsnProGluAsn 260  
 QY 217 TACTGGCAGACGCTTAGCACACACCGCTATATATACTTACGTACATAATCTCG-----ACAATT 270  
 DB 261 ArgTyrGlnThrAlaAlaGlu-----MetArgAlaAspLeuIleArgValHis 276  
 QY 271 GATGTGTCAGCGCTTGTTCATTACCATCCTATGCGACACGACAGATCCGTCGCTACTGCTT 330  
 DB 277 SerGlyGlnProGluAlaProLysValLeuThrAspAlaAspArgSerCysLeuLeu 296  
 QY 331 AGTGCTGCGCTACCCGCTCCACAAAGCGGAGGGGATTTGCGCTCGGACA 378  
 DB 297 SerSerGlyAla-----GlyAsnPheGlyValProArgThr 308  
 RESULT 11  
 RRPO.SHVX  
 ID RRPO.SHVX STANDARD; PRT; 1718 AA.  
 AC Q04575;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE RNA replication protein (194 kDa protein) (ORF 1) [Contains: RNA-  
 DE directed RNA polymerase (EC 2.7.7.48); Probable helicase].  
 OS Shalot virus X (SHVX).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Allexivirus.  
 OX NCBI\_TaxID=31770;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93019008; PubMed=1339468;  
 RA Kanyuka K.V., Vishnichenko V.K., Levay K.E., Kondrikov D.Y.,  
 RA Ryabov E.V., Zavrlev S.K.;  
 RT "Nucleotide sequence of shalot virus X RNA reveals a 5'-proximal  
 RT cistron closely related to those of potexviruses and a unique  
 RT arrangement of the 3'-proximal cistrons.";  
 RL J. Gen. Virol. 73:2553-2560(1992).  
 CC -1- FUNCTION: RNA-replication. The central part of this protein  
 CC possibly functions as an ATP-binding helicase.  
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +  
 CC [RNA](N).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M97264; AAA47787.1; -;  
 DR PIR; J01734; J01734.  
 DR InterPro; IPR005123; 2OG\_FeII\_Oxy.  
 DR InterPro; IPR001788; RNA\_dep\_RNAPol2.  
 DR Pfam; PF00978; RNA\_dep\_RNAPol2; 1.  
 DR Pfam; PF01443; Viral\_helicase1; 1.  
 DR Pfam; PF03171; 2OG\_FeII\_Oxy; 1.  
 KW ATP-binding; Helicase; RNA replication; RNA-directed RNA polymerase;  
 KW Transferase.  
 FT NP\_BIND 915 922 ATP (POTENTIAL).  
 SQ SEQUENCE 1718 AA; 194529 MW; 820FEAE1EB62415C CRC64;  
 Alignment Scores:  
 Pred. No.: 9.61 Length: 1718  
 Score: 73.50 Matches: 24  
 Percent Similarity: 44.30% Conservative: 11  
 Best Local Similarity: 30.38% Mismatches: 31



```
Query Match: 9.27% Indels: 13
DB: 1 Gaps: 3
US-09-856-221-1 (1-449) x RRPO_SHVX (1-1718)
QY 34 TATTTTTCATTGGAT-----AACGATTGGTCAGAACCC 66
|||||: |||
DB 826 TyrPheAsnLysAlaHisLeuArgLysAlaValLysThrAspMetSerProPro 845
QY 67 CGT-----TTAGACAGACCGCCAGTCNAACCATTCGTCATCATCATAGATATG 120
|||||: |||
DB 846 AlaArgProPheAspLeuAlaThrSerLeuHisArgGlyIleYrThrHisLysIle 865
QY 121 CGGCAACTGGCTCAACGCGCGCCCTTCCGCCGAAACGTAAGTAAATTCGTTAACCGCT 180
|||||: |||
DB 866 Asp-----AsnArgAlaThrAlaTyrMetSerAspValLysAsnAsnLeuThrGly 883
QY 181 TTGTTCTCTCCTCAGATAACAAACAACTCCAAAGTTACTGCGCAGACGTTAGCACAA 237
|||||: |||
DB 884 LeuValLeuProLysLeuAspArgAspLeuSerSerIrpValAlaLeuAlaGly 902
RESULT 12
YJJP_ECOLI
ID YJJP_ECOLI STANDARD; PRT; 256 AA.
AC P39402;
DT 01-FEB-1995 (Rel. 31, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Hypothetical protein YJJP.
GN YJJP OR B4364.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / WC1655;
RX MEDLINE=95334362; PubMed=7610040;
RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,
RA Blattner F.R.;
RT "Analysis of the Escherichia coli genome VI: DNA sequence of the
RT region from 92.8 through 100 minutes.";
RL Nucleic Acids Res. 23:2105-2119(1995).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: STRONG, TO H.INFLUENZAE H10108.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; U14003; AAA97263.1; ALT_INIT.
DR EMBL; AE000507; AAC77320.1; ALT_INIT.
DR Ecogene; EG12592; Yjip.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 118 138 POTENTIAL.
FT TRANSMEM 169 189 POTENTIAL.
FT TRANSMEM 195 215 POTENTIAL.
FT TRANSMEM 230 250 POTENTIAL..
SQ SEQUENCE 256 AA; 28001 MW; FC/FBE8ED0BE340A CRC64;
Alignment Scores:
Pred. NO.: 8.66 Length: 256
Score: 73.00 Matches: 26
Percent Similarity: 42.16% Conservative: 17
Best Local Similarity: 25.49% Mismatches: 35
Query Match: 9.21% Indels: 24
DB: 1 Gaps: 3
US-09-856-221-1 (1-449) x YJJP_ECOLI (1-256)
```

```
QY 169 TCGTTAACCGCTTTGTTCTCTT-----CCTCAGATA 198
|||||: |||
DB 148 SerThrAlaMetTyrIleArgGlnLeuAlaGlnArgHisLeuHisProGlnIle 167
QY 199 AACAAAAAAGTCTCAAGTTACTGGCAGACGTTAGCACCAACGCTATATAAATTACGTCAT 258
|||||: |||
DB 168 AsnPheCysLeuThrAlaPheAlaThr----- 177
QY 259 AATCTGACAAATTCAGTCAGCGCTTTCATTCATTCATTCATTCATTCATTCATTCATTC 318
|||||: |||
DB 178 -----ThrIleSerGlyLeuGlnLeuProThrPheSerAsn-----ThrPro 193
QY 319 TCGCTACTCTAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 378
|||||: |||
DB 194 ThrIleAlaMetAlaAlaSerValLeuLeuValProGlyPheProLeuIleAsnAla 213
QY 379 GTAATGCCGATGATCCGCTTTTCCGATTTATTCGAAATGCCCAAGTGGGAGTGACCCAA 438
|||||: |||
DB 214 ValAlaAspMetPheLysGlyHisIleAsnThrGlyLeuAlaArgTrpAlaIleAlaSer 233
QY 439 CTGATA 444
DB 234 LeuLeu 235
RESULT 13
E2F_DROME
ID E2F_DROME STANDARD; PRT; 805 AA.
AC Q27368; O77035;
DT 01-NOV-1997 (Rel. 35, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Transcription factor E2f (de2f).
GN E2F OR E2F1 OR CG6376
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Eye imaginal disk;
RX MEDLINE=94294381; PubMed=8022787;
RA Dynlacht B.D., Brook A., Dembski M., Yenush L., Dyson N.;
RT "DNA-binding and trans-activation properties of Drosophila E2F and DP
RT proteins.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:6359-6363(1994).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94158833; PubMed=8114698;
RA Ohtani K., Nevins J.R.;
RT "Functional properties of a Drosophila homolog of the E2F1 gene.";
RL Mol. Cell. Biol. 14:1603-1612(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Oregon-R;
RX MEDLINE=99077992; PubMed=9858578;
RA Sasaki T., Sawado T., Yamaguchi M., Shinomiya T.;
RT "Specification of regions of DNA replication initiation during
RT embryogenesis in the 65-kilobase DNAPolalpha-de2f locus of Drosophila
RT melanogaster.";
RL Mol. Cell. Biol. 19:547-555(1999).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
```

RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahle C., Davenport L.B., Davies P.,  
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Fostler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jatalil M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-P., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
RT "The genome sequence of *Drosophila melanogaster*.";  
RL Science 287:2185-2195(2000).  
RN [5]  
RN FUNCTION.  
RX MEDLINE=97415323; PubMed=9271122;  
RA Royzman I., Whitaker A.J., Orr-Weaver T.L.;  
RT "Mutations in *Drosophila* DP and E2F distinguish Gl-S progression from  
RT an associated transcriptional program.";  
RL Genes Dev. 11:1999-2011(1997).  
RN [6]  
RN FUNCTION.  
RX MEDLINE=96078671; PubMed=9418862;  
RA D'Antonio R.J., Bonnette P.C., O'Farrell P.H.;  
RT "Mutations of the *Drosophila* dDP, dE2F, and cyclin E genes reveal  
RT distinct roles for the E2F-dp transcription factor and cyclin E  
RT during the Gl-S transition.";  
RL Mol. Cell. Biol. 18:141-151(1998).  
RN [7]  
RN TISSUE SPECIFICITY, AND DEVELOPMENTAL STAGE.  
RX MEDLINE=96087653; PubMed=8537434;  
RA Hao X.F., Alphey L., Bandara L.R., Lam E.W., Glover D.,  
RA La Thangue N.B.;  
RT "Functional conservation of the cell cycle-regulating transcription  
RT factor DRP1/E2F and its pathway of control in *Drosophila*  
RT melanogaster.";  
RL J. Cell Sci. 108:2945-2954(1995).  
CC -!- FUNCTION: Transcriptional activator that binds to E2F sites.  
CC Required for wild-type growth in mitotic and polytene tissues,  
CC contributes to the expression of replication genes at the Gl-S  
CC transition and Cyclin E.  
CC -!- SUBUNIT: HETERODIMER OF E2F AND DP. COOPERATE TO GIVE SEQUENCE-  
CC SPECIFIC DNA BINDING AND OPTIMAL TRANS-ACTIVATION.  
CC -!- SUBCELLULAR LOCATION: Nuclear.  
CC -!- TISSUE SPECIFICITY: Segmentally repeated expression throughout  
CC early embryos is restricted to the ventral nerve cord in later  
CC embryos.  
CC -!- DEVELOPMENTAL STAGE: Throughout embryonic development.  
CC -!- SIMILARITY: BELONGS TO THE E2F/DP FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: X78421; CAA55186.1; -;  
DR EMBL: U10184; AAA19003.1; -;  
DR EMBL: AB011813; BAA32746.1; -;  
DR EMBL: AE003735; AAF55904.1; -;  
DR HSSP: Q16254; 1CF7.  
DR TRANSFAC: T01547; -;  
DR FLYBase: FBgn0011766; E2f.  
DR InterPro: IPR003316; E2F.TDP.  
DR Pfam: PF02319; E2F.TDP; 1.  
KW Transcription regulation; DNA-binding; Nuclear protein;  
FT Developmental protein; Activator.  
FT DNA\_BIND 253 318 POTENTIAL.  
FT DOMAIN 318 411 DIMERIZATION (POTENTIAL).  
FT DOMAIN 14 19 POLY-SER.  
FT DOMAIN 64 68 POLY-ASN.  
FT DOMAIN 115 125 POLY-ALA.  
FT DOMAIN 129 143 GLY-RICH.  
FT DOMAIN 245 249 POLY-SER.  
FT DOMAIN 519 573 GLY-RICH.  
FT DOMAIN 525 533 POLY-GLN.  
FT DOMAIN 594 601 POLY-ALA.  
FT DOMAIN 701 710 POLY-GLY.  
FT CONFLICT 127 127 H -> Q (IN REF. 1 AND 2).  
SQ SEQUENCE 805 AA; 87460 MW; BD49C813DDB2A77D CRC64;  
Alignment Scores:  
Pred. No.: 11.2 Length: 805  
Score: 72.50 Matches: 38  
Percent Similarity: 37.29% Conservative: 28  
Best Local Similarity: 21.47% Mismatches: 70  
Query Match: 9.14% Indels: 41  
DB: 1 Gaps: 5  
US-09-856-221-1 (1-449) x E2F\_DROME (1-805)  
QY 34 TATTTTCATTGATACGATTGGTCAGAACCCGCTTTAGAGAACGCCGACGTCACAAACC 93  
Db 104 TyrosineSerLeuAspGlyHisThrGlyAlaValAlaThrAlaAlaAlaAla 123  
QY 94 ATTCTGTCATCATTCATCAGCAATAAATCGGCAA-----CTGCGTCAA 135  
Db 124 ThrAlaGlyHisThrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 143  
QY 136 CGCGCGGCC-----TTCGCGCGCAACGTCATCGCA 165  
Db 144 ArgLysAlaThrGlyLysSerAsnAspIleThrAsnTyrTyrLysValLysArgArgPro 163  
QY 166 AATTCGTTAACCGCTTGTCTCTCTCAGATAAACAACAAAGTTCAGTTCGCGAG 225  
Db 164 HisAlaValSerAspGluLeuHisProLysGlnAlaLysGlnSerAlaHisGln 183  
QY 226 ACGTACGACAAACG-----CTATAT 246  
Db 184 ThrValTyrGlnLysHisThrAlaSerSerAlaProGlnGlnLeuArgHisSerHis 203  
QY 247 AACTTACGTCATATCTGACAAATTCATTCAGCGTTCAGCGTTCATCCCATCTATCGACA 306  
Db 204 GlnLeuArgHisAspAlaAspAlaGluLeuAspGluValValGluArgValAlaLys 223  
QY 307 CCAGCAGAT-----CCGTCCTACTGTTAGTCTGCCGTC 342  
Db 224 ProAlaSerHisHisProPheSerLeuSerThrProGlnGlnLeuAlaAlaSerVal 243  
QY 343 ACCGCCCTCAACAGCGGAGGATTCCTCGGACAGTAATGCGGATG----- 390  
Db 244 AlaSerSerSerSerGlyAspArgAsnArgAlaAspThrSerLeuGlyLeuThr 263  
QY 391 TACGTTTTCCGATTTATTCGAAATCCCAAGTGGGAGTGGAGTACCCCACTG 441  
Db 264 LysLysPheValAspLeuLeuGlnGluSerProAspGlyValValAspLeu 280



Thu Jan 16 09:33:19 2003

CC -|- CATALYTIC ACTIVITY: Succinyl-CoA + dihydrolipoamide = CoA + S-  
 CC succinylidihydrolipoamide.  
 CC -|- COFACTOR: THE E2 COMPONENT CONTAINS ONE COVALENTLY-BOUND LIPOYL  
 CC COFACTOR.  
 CC -|- PATHWAY: Tricarboxylic acid cycle.  
 CC -|- SUBUNIT: FORMS A 24-POLYPEPTIDE STRUCTURAL CORE WITH OCTAHEDRAL  
 CC SYMMETRY.  
 CC -|- SUBCELLULAR LOCATION: Mitochondrial.  
 CC -|- SIMILARITY: BELONGS TO THE 2-OXOACID DEHYDROGENASE FAMILY.  
 CC -|- SIMILARITY: CONTAINS 1 LIPOYL-BINDING DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; D90401; BA14397.1; -  
 DR PIR; A41015; A41015.  
 DR HSP; P07016; IC4T.  
 DR InterPro; IPR001078; 2-oxoacid\_dh.  
 DR InterPro; IPR000089; Biotin\_lipoyl.  
 DR InterPro; IPR003016; Lipoyl.  
 DR Pfam; PF00198; 2-oxoacid\_dh; 1.  
 DR Pfam; PF00364; biotin\_lipoyl; 1.  
 DR PRODOM; PD001115; 2-oxoacid\_dh; 1.  
 DR PROSITE; PS00189; LIPOYL; 1.  
 DR Tricarboxylic acid cycle; Transferase; Acyltransferase; Lipoyl;  
 KW Mitochondrion; Transit peptide.  
 FT TRANSIT 1 56 MITOCHONDRION.  
 FT CHAIN 57 442 DIHYDROLIPOAMIDE SUCCINYLTRANSFERASE  
 FT COMPONENT OF 2-OXOGLUTARATE DEHYDROGENASE  
 FT COMPLEX.  
 FT BINDING 99 99 LIPOYL (POTENTIAL).  
 FT ACT\_SITE 413 413 POTENTIAL.  
 FT ACT\_SITE 417 417 POTENTIAL.  
 SQ SEQUENCE 442 AA; 47412 MW; 5109DB7F1DDFC2F6 CRC64;

## Alignment Scores:

Pred. No.:	11.8	Length:	442
Score:	72.00	Matches:	27
Percent Similarity:	40.40%	Conservative:	13
Best Local Similarity:	27.27%	Mismatches:	37
Query Match:	9.08%	Indels:	22
DB:	1	Gaps:	3

US-09-856-221-1 (1-449) x ODO2\_RAT (1-442)

QY	64	CCCCGTTT	AGAGAGCCGCGAGTCAACCATTCGTGATCATATATCAGCATAAATGCGG	123
Db	198	ProLeuAlaGluAlaGlyAlaAlaLysGlyLeuArgSerGluHisArgGluLysMetAsn	217	
QY	124	CAACTGCGTCAACCGCGGCTTGGCGGGAACGCTACTGCAAAATTCGTTAACCGCTTG	183	
Db	218	ArgMetArgGlnArgIleAlaGlnArgLeuLysGluAlaGlnAsn-----ThrCysAla	235	
QY	184	TTCTCTCTCAGATAAACAACAACTGCAAAAGTTACTGGCAGACGTTAGCACACACGC---	240	
Db	236	MetLeuThrThrPheAsnGluValAspMetSerAsnIleGlnGluMetArgAlaArgHis	255	
QY	241	-----CTATATACCTTACGTATATCTGACAAATGTGTGTCAGCCGTTGTCATTACCC	294	
Db	256	LysAspAlaPheLeuLysLysHisAsnLeuLysLeuGly-----LeuMetSerAlaPheValysAlaser	277	
QY	295	ATCTATGCGACACGACGAGATCCGTCGTTAGTGTGTCGCGTCACCGCCTCA	351	
Db	269	-----LeuMetSerAlaPheValysAlaser	277	

Search completed: January 15, 2003, 15:46:24  
 Job time : 17.6207 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: January 15, 2003, 15:44:13 ; Search time 42.2998 Seconds  
(without alignments)  
4374.264 Million cell updates/sec

Title: US-09-856-221-1  
Perfect score: 793  
Sequence: 1 tgtgcagcactcacttat.....gtgacccaactgatacagtt 449

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Xgapop 10.0, Xgapext 0.5  
Xgapop 6.0, Xgapext 7.0  
Delop 6.0, Delxext 7.0

Searched: 671580 seqs, 206047115 residues  
Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+n2p.model -DEV=xlp  
-O=/cgn2.1/USPTO.spool/US09856221/runat\_15012003\_153925\_12457/app\_query.fasta\_1.1948  
-DB-SPTREMBL\_21 -OFT=fastan -SUFFIX=rspt -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40 cdi -LIST=45  
-DOALIGN=200 -THRESHOLD=pct -THRESHOLD=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09856221 -CGN\_1\_171 -runat\_15012003\_153925\_12457 -NCPU=6 -ICPU=3  
-NO\_XLPHY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120  
-WARN\_TIMEOUT=30 -THRESHOLD=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :  
1: sp\_archaea:  
2: sp\_bacteria:  
3: sp\_fungi:  
4: sp\_human:  
5: sp\_invertebrate:  
6: sp\_mammal:  
7: sp\_mhc:  
8: sp\_organelle:  
9: sp\_phage:  
10: sp\_plant:  
11: sp\_rodent:  
12: sp\_virus:  
13: sp\_vertebrate:  
14: sp\_unclassified:  
15: sp\_rvirus:  
16: sp\_bacteriap:  
17: sp\_archaeap:

Pred. NO. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	684	86.3	2523	2 Q93RP0	Q93rp0 xenorhabdus

2	434.5	54.8	2516	2	Q9RN43	Q9rn43 photorhabdu
3	416.5	52.5	2376	2	Q9F923	Q9f923 serratia en
4	415	52.3	2504	2	Q85160	Q85160 photorhabdu
5	335	42.2	2538	2	Q93RN7	Q93rn7 xenorhabdus
6	227	28.6	1189	2	Q85152	Q85152 photorhabdu
7	208	26.2	1565	2	Q85156	Q85156 photorhabdu
8	164	20.7	1279	2	Q93RP4	Q93rp4 xenorhabdus
9	84.5	10.7	270	16	Q8YZP6	Q8yzp6 anabena sp
10	84.5	10.7	568	16	Q8Z9G0	Q8z9g0 chlamydia p
c 11	84	10.8	338	16	Q8U8D8	Q8u8d8 agrobacteri
12	81	10.2	407	5	Q8T9A8	Q8t9a8 drosophila
13	81	10.2	578	5	Q9VMT7	Q9vmt7 drosophila
14	80.5	10.2	1275	17	Q8ZTL5	Q8ztl5 pyrobaculum
15	79	10.0	323	16	Q8XS16	Q8xsl6 ralstonia s
16	78.5	9.9	662	2	Q9L4V1	Q9l4v1 streptomyce
17	78	9.8	183	8	Q8WKU9	Q8wk9 pardanthops
18	78	9.8	468	3	Q9URQ0	Q9urq0 candida alb
19	78	9.8	640	12	Q5E294	Q5e294 human herpe
c 20	77.5	9.9	307	16	Q06723	Q06723 bacillus su
21	77.5	9.8	1881	12	Q9DUN3	Q9dun3 vesicular e
c 22	77	9.9	509	17	Q8TWC3	Q8twc3 methanopyru
23	77	9.7	535	5	Q26628	Q26628 strongyloce
24	77	9.7	1944	5	Q9N6N2	Q9n6n2 leishmania
25	77	9.7	3201	5	Q9W0U2	Q9w0u2 drosophila
26	76.5	9.6	315	2	Q93SS3	Q93ss3 pleiomonas
27	76.5	9.6	1240	12	Q9DMH8	Q9dwh8 rat cytomeg
c 28	76	9.6	457	5	Q9BHD0	Q9bhd0 leishmania
29	76	9.7	599	10	Q94E84	Q94e84 oryza sativ
30	75.5	9.5	298	5	Q9SP55	Q9sp55 aedes aegy
31	75	9.5	260	11	Q9LYM5	Q9lym5 mus musculu
32	75	9.5	298	2	Q9RHL0	Q9rhl0 rhodospseudo
33	75	9.5	466	3	Q9C1T1	Q9c1t1 aspergillus
34	75	9.5	537	5	Q8SQS3	Q8sq3 enccephalito
c 35	75	9.5	635	10	Q9SH38	Q9sh38 arabidopsis
36	75	9.5	1014	11	Q9VHJ7	Q9vhj7 mus musculu
37	74.5	9.4	407	16	Q9T089	Q9t089 streptococc
38	74.5	9.4	452	2	Q59558	Q59558 streptococc
39	74.5	9.4	522	5	Q8T773	Q8t773 branchiostoc
40	74.5	9.4	670	5	Q9N9C4	Q9n9c4 leishmania
41	74.5	9.4	709	12	Q9Q0Q5	Q9q0q5 influenza a
42	74.5	9.4	715	12	Q9Q0Q4	Q9q0q4 influenza a
43	74.5	9.4	1437	5	Q44329	Q44329 hirudo medl
44	74	9.3	558	3	Q96TU2	Q96tu2 pleurotus o
45	74	9.3	1186	5	Q17786	Q17786 caenorhabdi

## ALIGNMENTS

RESULT 1  
Q93RP0 PRELIMINARY; PRT; 2523 AA.  
ID Q93RP0  
AC Q93RP0  
DT 01-DEC-2001 (TRENBLrel. 19, Created)  
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)  
DT 01-MAR-2002 (TRENBLrel. 20, Last annotation update)  
DE XPTAL protein.  
GN XPTAL  
OS Xenorhabdus nematophilus.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Xenorhabdus.  
OX NCBI\_TaxID=628;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=PMF1296;  
RX MEDLINE=21218513; PubMed=11319082;  
RA Morgan J.A.W., Sergeant M., Ellis D., Ousley M., Jarrett P.;  
RT "Sequence analysis of insecticidal genes from Xenorhabdus nematophilus  
pmf1296.";  
RL Appl. Environ. Microbiol. 67:2062-2069(2001).  
DR EMBL; AJ308438; CAC38401.1;  
DR InterPro; IPR000953; Chromo.  
DR SMART; SM00298; CHROMO; 1.  
SQ SEQUENCE 2523 AA; 286999 MW; 3159852E0655B5B1 CRC64;

Alignment Scores: 1.06e-68 Length: 2523  
 Pred. No.: 684.00 Matches: 134  
 Score: 684.00  
 Percent Similarity: 93.96% Conservative: 6  
 Best Local Similarity: 89.93% Mismatches: 1  
 Query Match: 86.25% Indels: 0  
 DB: 2 Gaps: 0

US-09-856-221-1 (1-449) x Q93RP0 (1-2523)

QY 2 GTGCGAGCTACCTATTGGCGGATAA-CCTATTATTTTCATTGGATAACGATTGCTGCA 60  
 DB 1903 ValGlnAlaLeuThrLeuLeuGlyAspGluProTyrPheSerLeuAspAsnSprpSer 1922  
 QY 61 GAACCCCGTTAGAGAACGCCAGTCAAAACATTGCTGATCATATTATCAGCATAAATG 120  
 DB 1923 GluProArgLeuGluGluAlaAlaSerGlnThrMetArgHisHisTyrGlnHisLysMet 1942  
 QY 121 CGGCAACTGGCTCAACGCCGCGCTTGGCGGAAAGTACTGCAAAATTCGTTAAACGCT 180  
 DB 1943 LeuGlnLeuArgGlnArgAlaAlaLeuProThrLysArgThrAlaAsnSerLeuThrAla 1962  
 QY 181 TTGTTCTCTCAGATAACAAAACTGCAAAATTTACTGGCAGAGCTTACGACAAGCC 240  
 DB 1963 LeuPheLeuProGlnGlnLeuAsnLysLeuGlnGlyTyrTrpGlnThrLeuThrGlnArg 1982  
 QY 241 CTATATACTTACGTCATAATCTGCAATTTGATGTCAGCGCTGTGTCATTACCCATCTAT 300  
 DB 1983 LeuTyrAsnLeuArgHisAsnLeuThrIleAspGlyGlnProLeuSerLeuSerLeuTyr 2002  
 QY 301 CGGACCCAGCAGTCCGCTCTACTGCTAGTGTGCTGCTCCCTACCGCTCACAGCGGA 360  
 DB 2003 AlaThrProAlaAspProSerMetLeuLeuSerAlaAlaIleThrAlaSerGlnGly 2022  
 QY 361 GGGGATTGGCTCGGACAGTAATGCGATGATGCGCTTTCGGATTATTCGGAATGCC 420  
 DB 2023 GlyAspLeuProHisAlaValMetProMetTyrArgPheProValIleLeuGluAsnAla 2042  
 QY 421 AAGTGGGAGTGACCCCACTGATACAG 447  
 DB 2043 LysTrpGlyValSerGlnLeuIleGln 2051

RESULT 2  
 Q9RN43 PRELIMINARY; PRT: 2516 AA.  
 AC Q9RN43;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE Toxin A (Toxin complex protein).  
 GN TCDA.  
 OS Photorhabdus luminescens (Xenorhabdus luminescens).  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Photorhabdus.  
 OX NCBI\_TaxID=29488;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-W-14;  
 RA Merlo D.J., Wegrich L.M., Roberts J.L., Petell J.K.;  
 RT "Expression of Photorhabdus luminescens toxA gene in maize confers  
 corn rootworm tolerance.";  
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-W14;  
 RX MEDLINE=21185117; PubMed=11286884;  
 RA Waterfield N.R., Bowen D.J., Petherston J.D., Perry R.D.,  
 RA Wrench-Constant R.H.;  
 RA "The tc genes of Photorhabdus: a growing family.";  
 RT Trends Microbiol. 9:185-191(2001).  
 RL EMBL: AF188483; AAF0542.1; -.  
 DR EMBL; AF346500; AAL18486.1; -.

DR InterPro: IPR001589; Actbind\_actnin.  
 DR PROSITE: PS00019; ACTININ\_1; UNKNOWN\_1.  
 SQ SEQUENCE 2516 AA; 282952 MW; 8250A0650B614B99 CRC64;

Alignment Scores: 2.63e-40 Length: 2516  
 Pred. No.: 434.50 Matches: 87  
 Score: 434.50  
 Percent Similarity: 72.50% Conservative: 20  
 Best Local Similarity: 58.78% Mismatches: 38  
 Query Match: 54.79% Indels: 3  
 DB: 2 Gaps: 1

US-09-856-221-1 (1-449) x Q9RN43 (1-2516)

QY 13 CACCTTATGGCGGATAACCTTATTTTCATTGGATAACGATTGCTGCAACCCCGTTTA 72  
 DB 1888 HisLeuLeuGlyAspLysProTyrLeuProLeuSerThrThrTrpSerAspProArgLeu 1907  
 QY 73 GAAGAACGCCGAGTCAAAACCATTCGTGATCATATCAGCATATAAAATGGGCAACTCGCT 132  
 DB 1908 AspArgAlaAlaAspIleThrThrGlnAsnAlaHisAspSerAlaIleValAlaLeuArg 1927  
 QY 133 CAACGCCGCGCTTCCGCGGAAA-----CGTACTGCAAAATTCGTTAAACGCTTTG 183  
 DB 1928 GlnAsnIleProThrProAlaProLeuSerLeuArgSerAlaAsnThrLeuThrAspLeu 1947  
 QY 184 TTCCTTCTCAGATAACAAAACTGCAAAATTTACTGGCAGAGCTTACGACAAGCCCTA 243  
 DB 1948 PheLeuProGlnIleAsnGluValMetMetAsnTyrTrpGlnThrLeuAlaGlnArgVal 1967  
 QY 244 TATAACTTACGTCATAAATGACAAATTCGATGTCAGCGCTTGCATTTACCCATCTATGCG 303  
 DB 1968 TyrAsnLeuArgHisAsnLeuSerIleAspGlyGlnProLeuTyrLeuProIleTyrAla 1987  
 QY 304 ACACGACGAGATCCGCTCGTACTGCTTAGTGTGCTGCTCCGCTACCGCTCACAGCGGAGG 363  
 DB 1988 ThrProAlaAspProLysAlaLeuLeuSerAlaAlaValAlaThrSerGlnGlyGly 2007  
 QY 364 GATTTCCTCGGACAGTATCCGATGTCGCTTTCGCGATTATTCGGAATGCCAAG 423  
 DB 2008 LysLeuProGlnSerPheMetSerLeuTrpArgPheProHisMetLeuGluAsnAlaArg 2027  
 QY 424 TGGGAGTGACCAACTGATACAG 447  
 DB 2028 GlyMetValSerGlnLeuThrGln 2035

RESULT 3  
 Q9F923 PRELIMINARY; PRT: 2376 AA.  
 AC Q9F923;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE SepA.  
 GN SEPA.  
 OS Serratia entomophila.  
 OC Plasmid PADAP.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Serratia.  
 OX NCBI\_TaxID=42906;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ALM02;  
 RX MEDLINE=20416224; PubMed=10960097;  
 RA Hurst M.R., Glare T.R., Jackson T.A., Ronson C.W.;  
 RT "Plasmid-located pathogenicity determinants of Serratia entomophila,  
 the causal agent of amber disease of grass grub, show similarity to  
 the insecticidal toxins of Photorhabdus luminescens.";  
 RL J. Bacteriol. 182:5127-5138(2000).  
 DR EMBL: AF135182; AAG09642.1; -.  
 KW Plasmid.  
 SQ SEQUENCE 2376 AA; 262649 MW; E418DACE22DB87BF CRC64;

Alignment Scores: 2,89e-38 Length: 2376  
 Pred. No.: 416.50 Matches: 88  
 Score: 72.97% Conservative: 20  
 Best Local Similarity: 59.46% Mismatches: 39  
 Query Match: 52.52% Indels: 2  
 DB: 1

US-09-856-221-1 (1-449) x Q9F923 (1-2376)

QY 2 GTGAGGAGCTACCTATTATGGCGGATAA-CCTTATTTTTCATTGGATACGATTGGTCA 60  
 DB 1756 ValGlnAlaLeuAsnLeuLeuGlyAspGluProTyrLysPheAspAlaAspIrpSer 1775  
 QY 61 GAACCCGTTTACAAAGCCGCCAGTCAAAACCATTCGTGATCATTCATACGATAAAATG 120  
 DB 1776 AlaLeuThrLeuGlyAspAlaAlaSerGluValThrArgArgAspTyrGlnGluAlaLeu 1795  
 QY 121 CGCAACTGCGTCAACGCGCGCTGCGCGGAAACGCTACTGCAAAATTCGTTAACCGCT 180  
 DB 1796 LeuAlaValArgLeuValProAlaProGluThrArgThrAlaAsnSerLeuThrAla 1815  
 QY 181 TGTCTCTCTTCAGATTAACAAAACTGCAAGTTACTGGCAGAGCTTAGCACAAACGC 240  
 DB 1816 LeuPheLeuProGlnGlnAsnGluValLeuLysGlyTyrTrpGlnThrLeuAlaGlnArg 1835  
 QY 241 CTATATAACTTACGTATATCTGACAATTGATGTCAGCGGTTGTTCATTCACCATCTAT 300  
 DB 1836 LeuHisAsnLeuArgHisAsnLeuSerLeuAspGlyGlnProLeuSerLeuSerValTyr 1855  
 QY 301 GCGACACGAGATCCGTCGCTAGCTTACTGCTGCGTCAGCGCTCACAGGCGGA 360  
 DB 1856 AlaThrProSerGluProSerAlaLeuGlnSerAlaValAlaValAsnSerAlaGlnGlyAla 1875  
 QY 361 GGGGATTTGCTCGGACAGTAATGCCAGTGTACCGTTTCGGATTATTCGAAAATGCC 420  
 DB 1876 AlaAlaLeuProAlaAlaValMetProLeuTyrSerPheProValMetLeuGluAsnAla 1895  
 QY 421 AGTGGGAGTGACCAACTGATA 444  
 DB 1896 Arg---GlyMetValSerLeuLeu 1902

RESULT 4  
 O85160 PRELIMINARY; PRT: 2504 AA.  
 AC O85160:  
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)  
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE Insecticidal toxin complex protein TcBa.  
 GN TCBA.  
 OS Photobacterium luminescens (Xenorhabdus luminescens).  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Photobacterium.  
 OC NCBI\_TaxID=29488;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=W-14;  
 RX MEDLINE=98306208; PubMed=9641921;  
 RA Bowen D., Rocheleau T.A., Blackburn M., Andreev O., Golubeva E.,  
 Bhartiya R., fireCh-Constant R.H.;  
 RT "Insecticidal toxins from the bacterium Photobacterium luminescens.";  
 RL Science 280:2129-2132(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=W14;  
 RX MEDLINE=21185117; PubMed=11286884;  
 RA Waterfield N.R., Bowen D.J., Fetherston J.D., Perry R.D.,  
 fireCh-Constant R.H.;  
 RA "The tc genes of Photobacterium: a growing family.";  
 RT Trends Microbiol. 9:185-191(2001).  
 DR EMBL; AF047457; AAC38627.1; -;  
 DR EMBL; AF346498; AAL18460.1; -;

DR InterPro: IPR000566; Lipoclin\_cytFABP.  
 DR PROSITE; PS00213; LIPOCALIN; UNKNOWN.L.  
 SQ SEQUENCE 2504 AA; 280654 MW; 8F29B1693D1047CE CRC64;

Alignment Scores: 4,35e-38 Length: 2504  
 Pred. No.: 415.00 Matches: 90  
 Score: 71.33% Conservative: 17  
 Best Local Similarity: 60.00% Mismatches: 41  
 Query Match: 52.33% Indels: 3  
 DB: 1

US-09-856-221-1 (1-449) x O85160 (1-2504)

QY 5 CAGGACACTCACCTTATTTGGCGGATAA-CCTTATTTTTCATTGGATACGATTGGTCAGAA 63  
 DB 1880 GlnAlaLeuAsnLeuLeuGlyAspGluProGlnValMetLeuSerThrThrIrpAlaAsn 1899  
 QY 64 CCGCGTTTAGAAGAACGCCAGTCAAAACCATTCGTGATCATTCATCAGCATATAAATGCGG 123  
 DB 1900 ProThrLeuGlyAsnAlaAlaSerLysThrThrGlnGlnValArgGlnGlnValLeuThr 1919  
 QY 124 CAACCTGCGT-----CAACGCGCGGCTTGGCGCGAAACCTACTGCAAAATTCGTTAAC 177  
 DB 1920 GlnLeuArgLeuAsnSerArgValLysThrProLeuLeuGlyThrAlaAsnSerLeuThr 1939  
 QY 178 GCTTCTCTCTTCCCTCAGATAAACAACAACTGCAAAAGTTACTGGCAGAGCTTAGCACAA 237  
 DB 1940 AlaLeuPheLeuProGlnGlnAsnSerLysLeuLysGlyTyrIrpArgThrLeuAlaGln 1959  
 QY 238 CGCTATATAACTTACGTATATCTGACAATTGATGTCAGCGGTTGTTCATTCACCATC 297  
 DB 1960 ArgMetPheAsnLeuArgHisAsnLeuSerLeuAspGlyGlnProLeuSerLeuProLeu 1979  
 QY 298 TATGCGACACGACGAGATCCGTCGCTAGCTTACTGCTGCTGCTGCTGCTGCTGCTGCTG 357  
 DB 1980 TyrAlaLysProAlaAspProLysAlaLeuLeuSerAlaAlaValSerAlaSerGlnGly 1999  
 QY 358 GGAGGGGATTTGCTCGGACAGTAATGCCAGTGTACCGTTTACCGTTTTCGGATTATTCG 417  
 DB 2000 GlyAlaAspLeuProLysAlaProLeuThrIleHisArgPheProGlnMetLeuGluGly 2019  
 QY 418 GCCAAGTGGGAGTGACCCCACTGATACAG 447  
 DB 2020 AlaArgGlyLeuValAsnGlnLeuLeuGln 2029

RESULT 5  
 O93RN7 PRELIMINARY; PRT: 2538 AA.  
 AC O93RN7:  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE xptA2 protein.  
 GN xptA2.  
 OS Xenorhabdus nematophilus.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Xenorhabdus.  
 OC NCBI\_TaxID=628;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=PMF1296;  
 RX MEDLINE=21218513; PubMed=11319082;  
 RA Morgan J.A.W., Sergeant M., Ellis D., Ousley M., Jarrett P.;  
 RT "Sequence analysis of insecticidal genes from Xenorhabdus nematophilus  
 pmf1296.";  
 RL Appl. Environ. Microbiol. 67:2062-2069(2001).  
 DR EMBL; AJ308438; CAC38404.1; -;  
 SQ SEQUENCE 2538 AA; 283995 MW; 3A7A98CF54FD756E CRC64;

Alignment Scores: 5,55e-29 Length: 2538  
 Pred. No.: 335.00 Matches: 76  
 Score:







Alignment Scores:  
 Pred. No.: 1.01 Length: 270  
 Score: 84.50 Matches: 40  
 Percent Similarity: 46.96% Conservative: 14  
 Best Local Similarity: 34.78% Mismatches: 34  
 Query Match: 10.66% Indels: 27  
 DB: 16 Gaps: 8

US-09-856-221-1 (1-449) x Q8Y2P6 (1-270)

QY 91 ACATTCGTGATCATATATAGCAATAAATGCGGCAATGCGTCAACGGCGGCGCTTGCGG 150  
 Db 3 ThrileSerAspHisPheGln---SerLeuArgGlnArgGlnGlnCysAlaLeuilePro 21  
 QY 151 GCGAATCTACTGCAATTCGTTACCGCTTGTTCCTCTCAGATTAACAAAAAATCG 210  
 Db 22 -----phelleThrAlaGlyAspProAspLeu 30  
 QY 211 CAAGTTACTGCGAGAGCTTAGCACACGCCCTATATACTTA---CGTCATATCTGACA 267  
 Db 31 -----GlnThrThrAlaGluAlaLeuArgilleLeuAspArgHisGly----- 44  
 QY 268 ATTGATGTCAGCGGTGTTCATTACCATCTATGCGACACCA---GCAGATCGGTCGGTA 324  
 Db 45 AlaAspPheilleGluLeuGlyValPro---TyrSerAspProLeuAlaAspGlyProVal 63  
 QY 325 CTGCTTAGTCTGCGCTACCGCTCACAAAGCGGAGGGGATTTCCTCGG----- 375  
 Db 64 IleGlnAlaAlaAlaThrArgAlaLeuGlnArgGlyThrLysLeuGluGlnValLeuGlu 83  
 QY 376 -----ACAGTAATGCGGATGATACCGTTCCTTCGGATTATTCGTG 411  
 Db 84 MetLeuGluThrValilleProSerLeuGlnAlaProilleLeu 98

RESULT 10  
 Q929G0  
 ID Q929G0 PRELIMINARY; PRT; 568 AA.  
 AC Q929G0  
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)  
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Putative OMP protein.  
 GN CPN0021 OR CPJ0021 OR CP0755.  
 OS Chlamydia pneumoniae (Chlamydophila pneumoniae).  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydophila.  
 OX NCBI\_TaxID=83558;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CWL029;  
 RX MEDLINE=99208606; PubMed=10192388;  
 RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,  
 RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;  
 RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";  
 RL Nat. Genet. 21:385-389(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-AR39;  
 RX MEDLINE=20150255; PubMed=10684935;  
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heideberg J.F.,  
 RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,  
 RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,  
 RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,  
 RA Eisen J., Fraser C.M.;  
 RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia  
 RT pneumoniae AR39.";  
 RL Nucleic Acids Res. 28:1397-1406(2000).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-J138;  
 RX MEDLINE=20330349; PubMed=10871362;  
 RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,  
 RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;

"Comparison of whole genome sequences of Chlamydia pneumoniae J138  
 RT from Japan and CWL029 from USA.";  
 RL Nucleic Acids Res. 28:2311-2314(2000).  
 DR EMBL; AE001587; AAD18174.1;  
 DR EMBL; AE002234; AAF38558.1;  
 DR EMBL; AF002545; BAA98233.1;  
 DR TIGR; CP0755;  
 KW Complete proteome.  
 SQ SEQUENCE 568 AA; 63553 MW; 3358FCB5BC482E80 CRC64;

Alignment Scores:  
 Pred. No.: 1.22 Length: 568  
 Score: 84.50 Matches: 40  
 Percent Similarity: 42.66% Conservative: 21  
 Best Local Similarity: 27.97% Mismatches: 55  
 Query Match: 10.66% Indels: 27  
 DB: 16 Gaps: 6

US-09-856-221-1 (1-449) x Q929G0 (1-568)

QY 58 TCAGAACCCCGTTTGAAGAAGCC-----GCCAGTCAAAACCAATTCGTGAT----- 102  
 Db 232 AlaserProGlnAspGlnGluAlaileLeuTyraAlaLeuGlyLysLeuLysAspGlyGln 251  
 QY 103 CATTATCAGCATAAATGCGCAACTGCTCAACGGCGGCTTCGCGCGAAACGTACT 162  
 Db 252 SerTyTyraAsnIleLysLysGlnLeuLysProAspValaspaValThrLeuAlaala 271  
 QY 163 GCAATTCGTTAACCGCTTG-----TTCCTTCCTCAGATTAACAAAAA 207  
 Db 272 AlaglnAlaLeuAlaLeuGlyLysGlnGluAspAlaLeuProValilleLysLys--- 290  
 QY 208 CTGCAAAAGTTACTGGCAGACGCTTAGCACAAACGC-----CTATATACTTACGTAT 258  
 Db 291 -----GlnAlaLeuGluArgProArgAlaLeuTyraAlaLeuArgHis 305  
 QY 259 ANTCTGCAATTTGATGTCAGCGTGTCTATTACCATTCATCCCATCTATGCCAGCAGATCCG 318  
 Db 306 LeuProSerGluilleGlylleProilleAlaLeuProillePheLeuLysThrLysAsnSer 325  
 QY 319 TCGGTACTGTTAGTGTCTGCGCTCACCGCTCACAAAGCGGAGGGGATTCGCT----- 372  
 Db 326 GluAlaLysLeuAsnValaLeuAlaLeuLeuGlyLysCysAspThrProLysLeu 345  
 QY 373 -----CGGACAGTAATCGCATGTACCGTTTTCGATTTATTCCTGGAA 414  
 Db 346 LeuGluTyrrilleThrGluArgLeuValGlnProHisTyraGlnThrLeuAlaLeuSer 365  
 QY 415 AATGCCAAG 423  
 Db 366 PheSerLys 368

RESULT 11  
 Q8U8D8  
 ID Q8U8D8 PRELIMINARY; PRT; 338 AA.  
 AC Q8U8D8  
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 DE ABC transporter, substrate binding protein.  
 GN ATU4154 OR AGR\_L\_1399.  
 OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Rhizobiaceae; Rhizobium.  
 OX NCBI\_TaxID=176299;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21608550; PubMed=11743193;  
 RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,  
 RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,  
 RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,  
 RA Kutayavin T., Levy R., Li M.-J., McClelland E., Palmieri A.,

RA Raymond C., Rouse G., Saenphimachak C., Wu Z., Romero P., Gordon D.,  
 RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,  
 RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,  
 RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,  
 RA Nester E.W.;  
 RT "The genome of the natural genetic engineer Agrobacterium tumefaciens  
 RT C58.";  
 RL Science 294:2317-2323(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=21608551; PubMed=11743194;  
 RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M., Mullin L.,  
 RA Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,  
 RA Houmel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,  
 RA Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,  
 RA Flanagan C., Crowell C., Gursun J., Lomo C., sear C., Strub G.,  
 RA Cielo C., Slater S.;  
 RA "Genome sequence of the plant pathogen and biotechnology agent  
 RT Agrobacterium tumefaciens C58.";  
 RL Science 294:2323-2328(2001).  
 DR EMBL; AE009345; AAL44954.1; ALT\_INIT.  
 DR EMBL; AE008268; AAK89280.1; -;  
 KW Complete proteome.  
 SQ SEQUENCE 338 AA; 35571 MW; 879AA0BEFDD3C34F CRC64;

Alignment Scores:  
 Pred. No.: 1,22 Length: 338  
 Score: 84.00 Matches: 42  
 Percent Similarity: 43.70% Conservative: 17  
 Best Local Similarity: 31.11% Mismatches: 53  
 Query Match: 10.76% Indels: 23  
 DB: 16 Gaps: 8

US-09-856-221-1 (1-449) x Q8UBD8 (1-338)

QY 443 ATCAAGTGGGTCACT-----CCCCACTTGGCATTTTCAGATAATCGGAAA 396  
 Db 83 ValSerTrpValGluPheThrAlaGlyProProLeuValGluAlaLeuAsnValGlySer 102  
 QY 395 CGGTACATCGCATCTACTGTCGAGCAAAATCCCTCCG----- 357  
 Db 103 IleAsnValGlyTrpThr-----GlyAspAlaProProIlePheGlyGlnAlaGly 120  
 QY 356 CTTGTGAGCGGTGACGGCAGCAGCACTAGCAGGATCTGCTGGTGGTCCGATAG 297  
 Db 121 SerAlaIleValTyValAlaAlaLeuProSer---AsnGlyLysGlyGluAlaIlePhe 139  
 QY 296 ATGGGTAATGACACGGCTGACCATCAATGTGCAGATTATGACGTAAGTTATAGCGT 237  
 Db 140 ThrLysProGluSerGlyIleLysSerVal-AlaAspLeuLysGlyLysValGlyVa 159  
 QY 236 TGTGCTA---ACGTCTGCCAGTAACTTTGCAGTTTTCGTTTATCTGAGGAAGAACAA 180  
 Db 159 IGlyLysGlyThrSerAlaHisAsnLeuLeuValAlaAlaLeuGluLysAsnGlyLeuLy 179  
 QY 179 GCGGTTACGAATTGCG-----TACGTTTCGCCGCAAGCGCGCGTTCAGCGAGT 126  
 Db 179 s---LeuSerAspIleAspValThrTyLeuSerProAla-----AspAlaAl 194  
 QY 125 TGCCGCAATTTCGTGATAATGATCAGGAATGTTTGACTCG 83  
 Db 194 aAlaAlaPheAlaSerAspLysIleAspAlaTrpAlaValTrp 208

RESULT 12  
 Q8T9A8 PRELIMINARY; PRT; 407 AA.  
 AC Q8T9A8;  
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE S010272P.  
 GN MSP-300.  
 OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,  
 RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,  
 RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,  
 RA Nunoo J., Pacleb J., Paragas V., Park S., Phouenavong S., Wan K.,  
 RA Yu C., Lewis S.E., Rubin G.M., Celisner S.;  
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AY069852; AAL39997.1; -;  
 SQ SEQUENCE 407 AA; 47289 MW; 9E76C03821C0CD40 CRC64;

Alignment Scores:  
 Pred. No.: 2,8 Length: 407  
 Score: 81.00 Matches: 27  
 Percent Similarity: 41.28% Conservative: 18  
 Best Local Similarity: 24.77% Mismatches: 40  
 Query Match: 10.21% Indels: 24  
 DB: 3 Gaps: 3

US-09-856-221-1 (1-449) x Q8T9A8 (1-407)

QY 61 GAACCCCGTTTGAAGAA----- 78  
 Db 294 AspProArgLeuSerGluAsnValAspAspLeuValArgValLeuAspGlyLeuGlu 313  
 QY 79 -----GCCGCGAGTCAACACCATTCGT-----GATCATTTATCAG 111  
 Db 314 LeuValLysAlaLysSerGlnSerLeuGluGlnThrLeuAlaGlnIleAspValTyGln 333  
 QY 112 CATATAATGCGGCAACTGCTCAACGCGCGCTTGCCTGCGGAAAGCTACTGCAAAATTCG 171  
 Db 334 GlnGlnMetGlnSerLeuArgGlnArgIleGlnGlnGlnGlnGlnLeuArgLeuVal 353  
 QY 172 TTAAACGCTTTGTTCTCTCCCTCAGATAACAAACAACTGCAAAAGTTACTGGCAGAGCTTA 231  
 Db 354 MetAlaProThrTyLeuProHisAspArgGluArgAlaLeuAlaGlnGlnAlaCys 373  
 QY 232 GCACAAAGCGCTATATACCTACGTACATAATCTGCAATTCGATGTCAGCGGTTGTCATTA 291  
 Db 374 ArgGluArgValLysAsnLeuHisSerLysIleThrAlaArgAsnGluArgIleLysLeu 393  
 QY 292 CCATCTAT---CCGACACCCACAGAT 315  
 Db 394 LeuIleHisArgGlyThrProAspAsp 402

RESULT 13  
 Q9VMT7 PRELIMINARY; PRT; 578 AA.  
 AC Q9VMT7;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE CG18251 protein.  
 GN MSP-300 OR CG18251 OR CG18252.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY.  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foster C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lin Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei R., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yen R., Zhong F.N., Zhong W., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RA "The genome sequence of Drosophila melanogaster";  
 RL Science 287:2185-2195(2000).  
 DR EMBL; AE003608; AAF52224.1; -  
 DR FlyBase; FBgn0010070; Msp-300.  
 DR InterPro; IPR002017; Spectrin.  
 DR SMART; SM00150; SPEC; 1.  
 SQ SEQUENCE 578 AA; 67145 MW; 27B0BAD7224A685D CRC64;

Alignment Scores:  
 Pred. No.: 3.06 Length: 578  
 Score: 81.00 Matches: 27  
 Percent Similarity: 41.28% Conservative: 18  
 Best Local Similarity: 24.77% Mismatches: 40  
 Query Match: 10.21% Indels: 24  
 Db: 5 Gaps: 3

US-09-856-221-1 (1-449) x 09VMT7 (1-578)

Qy 61 GAACCCCGTTTAGAAGAA----- 78  
 Db 465 AspProArgLeuSerGluAsnValAspAspLeuValArgValLeuAspGlyLeuGlu 484  
 Qy 79 -----GCCCCAGTCAACACATTCGT-----GATCATTTATCAG 111  
 Db 485 LeuValLysAlaLysSerGlnSerLeuGlnThrLeuAlaGlnLeuValTyrgln 504  
 Qy 112 CATAAATGCGCAACTGCGTCAACGCGCGCTTGCGCGGAAACGCTACTGCAAAATTCG 171  
 Db 505 GlnGlnMetGlnSerLeuArgGlnArgIleIleGlnGluGlnGlnLeuVal 524  
 Qy 172 ITAACCGCTTGTCTCTCCAGATAAACAACAACTGCAAGTACTGGCAGCGTTA 231  
 Db 525 MetAlaProThrTyrlLeuProHisAspArgGluArgAlaLeuAlaGluGlnAlaCys 544  
 Qy 232 GCACACGCTATATACTACGTCAATATCTGCAATATGTCAGCGCTGTTCATTA 291  
 Db 545 ArgGluArgValLysAsnLeuHisSerLysIleThrAlaArgAsnGluArgIleLysLeu 564  
 Qy 292 CCCATCTAT--GCGACACACGACAT 315  
 Db 565 LeuIleHisArgGlyThrProAspasp 573

RESULT 14

Q8ZTL5  
 ID Q8ZTL5 PRELIMINARY; PRT; 1275 AA.  
 AC Q8ZTL5;  
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE ATP-dependant helicase, probable.  
 GN PAE3198  
 OS Pyrobaculum aerophilum.  
 OC Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;  
 OC Thermoproteaceae; Pyrobaculum.  
 ON NCBI\_TaxID=13773;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=IM2 / ATCC 51768 / DSM 7523;  
 EX PubMed=11792869;  
 RA Fitz-Gibbon S.T., Ladner H., Kim U.-J., Stetter K.O., Simon M.I.,  
 RA Miller J.H.;  
 RT "Genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum  
 RT aerophilum";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:984-989(2002).  
 DR EMBL; AF009916; AAL64745.1; -  
 DR InterPro; IPR001410; DEAD.  
 DR InterPro; IPR001650; Helicase\_C.  
 DR Pfam; PF00270; DEAD; 1.  
 DR Pfam; PF00271; Helicase\_C; 1.  
 DR SMART; SM00487; DEXDC; 1.  
 DR SMART; SM00490; HELICC; 1.  
 KW Helicase; Complete proteome.  
 SQ SEQUENCE 1275 AA; 143418 MW; 866C3439310EC090 CRC64;

Alignment Scores:  
 Pred. No.: 4.27 Length: 1275  
 Score: 80.50 Matches: 25  
 Percent Similarity: 44.44% Conservative: 19  
 Best Local Similarity: 25.25% Mismatches: 36  
 Query Match: 10.15% Indels: 19  
 Db: 2 Gaps: 2

US-09-856-221-1 (1-449) x Q8ZTL5 (1-1275)

Qy 73 GAAGAAGCCGCGAGTCAACATTCGTGATCATATATCAGCATATAAATCGGCACTGGT 132  
 Db 570 AspAspAlaLeuSerSerSerIleIleValThrAsnProTyrlLysLeuSerSerMetLeu 589  
 Qy 133 CAACGCGCGCTTGCGCGGAAACGCTACTGCTTAACTGCTTAACTGCTTCTCTCT 192  
 Db 590 MetArgSerAlaAspSerAlaLysLysLeuAlaAsnSerLeuAlaLeuValIleAsp 609  
 Qy 193 CAGATAAACAACAACTGCAA-----AGTTACTGCGAGCAGCTTAGCACACGC 240  
 Db 610 GluValHisThrMetLeuGluProLysHisLeuAspPheThrAlaLeuLeuHisArg 629  
 Qy 241 CTATATTAACCTACGTCAATATCTGACAAATGATGTCAGCGCTGTTCATTCATCAT 300  
 Db 630 LeuTyrlLeuGluGlyAsp----- 635  
 Qy 301 GCGACACACGAGATCCGCTGCTACTGCTTACTGCTGCTGCTGCTGCTGCTGCTGCT 357  
 Db 636 ---ValLysLysTyrlProAlaIleIleLeuSerSerAlaThrValThrSerSerGly 653

RESULT 15  
 Q8XS16  
 ID Q8XS16 PRELIMINARY; PRT; 323 AA.  
 AC Q8XS16;  
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Probable magnesium and cobalt transport transmembrane protein.  
 GN RSP0665 OR RS01784.  
 OS Ralstonia solanacearum (Pseudomonas solanacearum).  
 OG Plasmid megaplasmid.  
 OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;

OC Ralstonia.  
 OX NCBI\_TaxID=305;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GMI1000;  
 RX MEDLINE=21681879; PubMed=11823852;  
 RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,  
 RA Ariat M., Billault A., Brottier P., Camus J.C., Catolico L.,  
 RA Chandelier M., Choisme N., Claudel-Renard C., Cunnac S., Demange N.,  
 RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,  
 RA Siquier P., Thebault P., Whalen M., Wincker P., Levy M.,  
 RA Weissbach J., Boucher C.A.;  
 RT "Genome sequence of the plant pathogen Ralstonia solanacearum";  
 RL Nature 415:497-502(2002).  
 DR EMBL; AL646080; CAD17816.1; -.  
 DR InterPro; IPR002523; CoRA.  
 DR Pfam; PF01544; CoRA; 1.  
 KW Plasmid; Complete proteome.  
 SQ SEQUENCE 323 AA; 36251 MW; DE264559E199DD61 CRC64;

Alignment Scores:  
 Pred. No.: 4.45 Length: 323  
 Score: 79.00 Matches: 46  
 Percent Similarity: 37.57% Conservative: 22  
 Best Local Similarity: 25.41% Mismatches: 66  
 Query Match: 9.96% Indels: 47  
 DB: 16 Gaps: 7

US-09-856-221-1 (1-449) x Q8XS16 (1-323)

QY 3 TGCAGGCACCTACCTTAT-----TGCGCGATAACC 32  
 DB 126 CysGluHisValProHisLeuLeuAlaLysGlyAlaProPheAlaLeuTyrAlaValMet 145  
 QY 33 TTATTTTTCATTGGATACGATTGGTCAGAACCCCGTTAGAA-----GAAGCCGCCAG 86  
 DB 146 AspPheValVal-AspAsn-----TyrGlnProValLeuGluSerMetGluAlaLaph 163  
 QY 87 TCAACCACTTCGTGATCAT----- 105  
 DB 163 eAspAlaIleGluGlnLeuPheGlyAspAlaPheAspArgAlaAlaIleGluArgLe 183  
 QY 106 -TATCAGCATAAATCGGCAACTCGCTCAACGC---GCGCGCTTCCGCGCAACAGTAC 161  
 DB 183 uTyrThrLeuLysArgGlnLeuLeuArgLeuArgAsnAlaAlaLeuProValGluAspI 203  
 QY 162 TGCAAAATCGTTAACCGCTTTGTTCTTCCTCAGATACAAAAAACTGCAAGTTACTG 221  
 DB 203 eAlaGlyGlnLeuValArgLeuHisGluAspValValProLysGluLeuArgAlaTyrPh 223  
 QY 222 GCAGACGTTAGCACACGCTATATAACTTACGTCAATAATGCAATTCATGTCAGCC 281  
 DB 223 eArgAspValAlaAspHisAlaHisArgLeuValGlyAlaLeuAspValIleArgGluMe 243  
 QY 282 GTTGTCAATCCCATCTATGCCAGACAGCAGATCCGCTGCTAGTGTGCTGCCGT 341  
 DB 243 tLeuThrThrAlaIleSerValAsnValAlaLeuValSerValThrGlnAsnAspIleVa 263  
 QY 342 CACCGCTCAACAGCGGAGGG-----GATTTCCTCGGACATATCCGATGTA 392  
 DB 263 lLysArgLeuAlaGlyTrpGlyAlaIleLeuAlaIleProThrValValPheSerAsnTy 283  
 QY 393 C-----CGTTTCCGATTATTCT 410  
 DB 283 rGlyMetAsnPhelyGlyMetProGluLeuGluHisProAlaGlyTyrProIleValle 303  
 QY 411 G 411  
 DB 303 u 303

Search completed: January 15, 2003, 15:54:20  
 Job time : 53.2998 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: January 15, 2003, 15:39:33 ; Search time 14.2569 seconds  
(without alignments)  
2878.683 Million cell updates/sec

```

Title:          US-09-856-221-2
Perfect score:  265
Sequence:       1 caaacatttgtgattta.....gcagacgttagcacacgcc 154

```

Scoring table:

	BLOSUM62	
Xgapop	10.0	Xgapext 0.5
Ygapop	10.0	Ygapext 0.5
Fgapop	6.0	Fgapext 7.0
delop	6.0	Delext 7.0

Searched: 908470 seqs, 133250620 residues  
Total number of hits satisfying chosen parameters: 1816940

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

```

Command line parameters:
-MODEL=framet_n2p_model
-Q/-cpn2.1/USPTO_spool_DEV09856221/runat_15012003_153924_12438/app_query.fasta_1.1948
-DB=A_Geneseq.101002_QPWT-fastan_SUFFIX=rgg_MINMATCH=0.1_LOOPCL=0
-LOOPEXT=0_UNITS=bits_START=1_END=-1_MATRIX=blosum62_TRANS=human40_cdi
-LIST=45_DOCALIGN=200_THR_SCORE=pct_THR_MAX=100_THR_MIN=0_ALIGN=15
-MODE=LOCAL_OUTFWT=pct_NORM=ext_HEAPSIZE=500_MINLEN=0_MAXLEN=200000000
-USER=US09856221.ecgn_1.160.&runat_15012003_153924_12438_NCPU=6_ICPU=3
-NO_XLUPY_NO_MMAP_LARGEQUERY_NEG_SCORES=0_WAIT_LONGLOG_DEV_TIMEOUT=120
-WARN_TIMEOUT=30_THRPADS=1_XGAPOP=10_YGAPEXT=0.5_FGAPOP=6_FGAPEXT=7
-YGAPOP=10_YGAPEXT=0.5_DELOP=6_DELEXT=7

```

Database :

A\_Geneseq\_101002:\*

1:	/SID52/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT:*
2:	/SID52/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT:*
3:	/SID52/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT:*
4:	/SID52/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT:*
5:	/SID52/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT:*
6:	/SID52/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT:*
7:	/SID52/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT:*
8:	/SID52/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT:*
9:	/SID52/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT:*
10:	/SID52/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT:*
11:	/SID52/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT:*
12:	/SID52/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT:*
13:	/SID52/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT:*
14:	/SID52/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT:*
15:	/SID52/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT:*
16:	/SID52/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT:*
17:	/SID52/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT:*
18:	/SID52/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT:*
19:	/SID52/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT:*
20:	/SID52/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT:*
21:	/SID52/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT:*
22:	/SID52/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT:*
23:	/SID52/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	137	51.7	2376	22	AA977694	SepA protein encoded by <i>Photobacterium</i> lumin
2	126.5	47.7	2504	18	AAW167871	Toxin TcbA, encoded by <i>Photobacterium</i> tcba
3	126.5	47.7	2504	19	AAW365571	Modified Photobacterium toxin TcdA, encoded by <i>Photobacterium</i> tcda
4	126.5	47.7	2504	22	AAW26121	Modified Photobacterium toxin TcdA, encoded by <i>Photobacterium</i> tcda
5	126.5	47.7	2505	22	AAW26121	Modified Photobacterium toxin TcdA, encoded by <i>Photobacterium</i> tcda
6	118.5	44.7	2516	18	AAW17899	TcdA toxin-zein ER
7	118.5	44.7	2516	22	AAW36572	Photobacterium lumin
8	118.5	44.7	2516	22	AAW26209	Photobacterium lumin
9	118.5	44.7	2517	22	AAW32611	Modified Photobacterium toxin TcdA, encoded by <i>Photobacterium</i> tcda
10	118.5	44.7	2537	22	AAW32611	Modified Photobacterium toxin TcdA, encoded by <i>Photobacterium</i> tcda
11	115.5	43.6	2522	20	AAW33729	Photobacterium lumin
12	113	42.6	573	18	AAW18303	Photobacterium lumin
13	113	42.6	573	19	AAW36559	Toxin TcbAIII, encoded by <i>Photobacterium</i> tcbaIII
14	108	40.8	579	18	AAW18301	Photobacterium lumin
15	108	40.8	579	19	AAW36574	Toxin TcdAIII, encoded by <i>Photobacterium</i> tcdaIII
16	63.5	24.0	886	22	ABW71328	Drosophila melanogaster
17	63	23.8	364	23	AAU99816	Escherichia coli M
18	61	22.8	114	22	AAU56070	Propionibacterium
19	60.5	22.8	658	22	ABW61728	Drosophila melanogaster
20	59.5	22.5	54	22	AAU44965	Propionibacterium
21	59.5	22.5	66	22	AAU39885	Propionibacterium
22	59	22.1	118	22	AAO00632	Human polyptide
23	58.5	21.9	639	22	ABW65465	Drosophila melanogaster
24	58.5	21.9	676	22	ABW59964	Drosophila melanogaster
25	58.5	21.9	676	22	ABW66818	Drosophila melanogaster
26	57	21.3	125	21	AAW24930	Drosophila melanogaster
27	57	21.3	254	21	AAW24446	Human ORF2694
28	57	21.3	276	21	AAW42947	Human secreted protein
29	57	21.3	299	21	AAW23042	Human ORF2711
30	57	21.3	316	22	AAU27957	Human myeloid upregulation
31	57	21.3	318	22	ABG22136	Human contig polyptide
32	57	21.3	322	21	AAW15549	Novel human diagnostic
33	57	21.3	322	21	AAW19729	Human immune system
34	57	21.3	322	21	AAW99349	Human SECC Clone 4
35	57	21.3	322	22	AAU27785	Human PRO110 (UNQ)
36	57	21.3	322	22	AAU29109	Human full-length
37	57	21.3	322	22	AAW93665	Human PRO polyptide
38	57	21.3	322	22	AAW66098	Human polyptide, protein of the inv
39	56.5	21.2	584	22	AAW96296	Putative P. abyssal
40	56.5	21.2	951	16	AAW57504	Eph-related CEK6
41	56	21.1	355	22	AAU34438	E. coli cellular protein
42	56	21.1	355	22	AAG98406	Escherichia coli protein
43	56	21.1	446	21	AAG57299	Arabidopsis thaliana
44	56	21.1	446	21	AAW61050	Arabidopsis thaliana
45	56	21.1	483	21	AAG57298	Arabidopsis thaliana

## ALIGNMENTS

RESULT 1	
AAAY97694	
ID	AAAY97694 standard; Protein; 2376 AA.
XX	
XX	AAAY97694;
XX	
XX	19-JUN-2001 (first entry)
DT	
DE	SepA protein encoded by Serratia insecticidal protein complex gene.
XX	
XX	Insecticidal protein complex; amber disease; insect; Coleoptera;
KW	pesticide; SepA protein.
KW	
XX	
XX	Serratia sp.
OS	
XX	
PN	WO200116305-A2.
XX	
XX	08-MAR-2001.
PD	
XX	
XX	

PF 04-SEP-2000; 2000WO-NZ00174.  
 XX 02-SEP-1999; 99NZ-0337610.  
 PR (NZPA-) NEW ZEALAND PASTORAL AGRIC RES INST LTD.  
 XX Glare TR, Hurst MRH, Jackson TA;  
 XX WPI; 2001-169009/17.  
 DR N-PSDB; AAA91292.  
 XX  
 PT New nucleic acid encoding a polypeptide useful as a pesticide  
 PT especially for Coleoptera -  
 XX  
 PS Claim 24; Page 92-99; 109pp; English.  
 XX  
 CC This sequence represents the SepA protein encoded by the Serratia  
 CC insecticidal protein complex gene of the invention. The invention relates  
 CC to a gene encoding an insecticidal protein complex or a functional  
 CC fragment, a neutral mutation, or a homologue of the complex. The  
 CC polypeptides and nucleotides of the invention are used to induce amber  
 CC diseases or like conditions in insects, especially to those from the  
 CC order comprising Coleoptera, useful as a pesticide.  
 XX  
 SQ Sequence 2376 AA;  
 Alignment Scores:  
 Pred. No.: 7.23e-11 Length: 2376  
 Score: 137.00 Matches: 28  
 Percent Similarity: 66.67% Conservative: 6  
 Best Local Similarity: 54.90% Mismatches: 17  
 Query Match: 51.70% Indels: 0  
 DB: 22 Gaps: 0  
 US-09-856-221-2 (1-154) x AAY97694 (1-2376)  
 QY 1 CAACACCTTCGTGATCATTATACGATAAAATCGCGCACTCGCTCAACGGCGGCGCTTG 60  
 DB 1785 GluValThrArgArgAspTyrGlnGluAlaLeuLeuAlaValArgArgLeuValProAla 1804  
 QY 61 CGCGCAACCTACTCGTAATTCGTTAAACCGCTTTGTTCTCTCCTCAGATAAACAACAAA 120  
 DB 1805 ProGluThrArgThrAlaAsnSerLeuThrAlaLeuPheLeuProGlnGlnAsnGluVal 1824  
 QY 121 CTGCAAACTTACTGCGACAGCTTAGCACACGC 153  
 DB 1825 LeuLysGlyTyrTrpGlnThrLeuAlaGlnArg 1835  
 RESULT 2  
 AAW17871  
 ID AAW17871 standard; Protein; 2504 AA.  
 AC AAW17871;  
 XX 29-JAN-1998 (first entry)  
 XX  
 XX Photorhabdus luminescens insect toxin protein TcBA.  
 DE  
 XX Insecticide; insect; toxin; pest control; biological control;  
 KW Photorhabdus luminescens; TcBA; Southern corn rootworm;  
 KW Colorado potato beetle; Western corn rootworm; meal worm;  
 KW boll weevil; turf grub; Coleoptera; beet armyworm; black cutworm;  
 KW cabbage looper; codling moth; corn earworm; European corn borer;  
 KW tobacco hornworm; tobacco budworm; Lepidoptera; hymenoptera;  
 KW Diptera, Dictyoptera; Acarina; Homoptera.  
 XX  
 OS Photorhabdus luminescens strain W-14 (ATCC 55397).  
 XX  
 PN W09717432-A1.  
 XX  
 PD 15-MAY-1997.  
 XX  
 PF 06-NOV-1996; 96WO-US18003.

XX 28-AUG-1996; 96US-0705484.  
 PR 06-NOV-1995; 95US-0007255.  
 PR 28-FEB-1996; 96US-0608423.  
 XX  
 PA (WISC ) WISCONSIN ALUMNI RES FOUND.  
 XX  
 XX Blackburn MB, Bowen DJ, Ciche TA, Ensign JC, Fatig R;  
 PI French-Constant RH, Guo L, Hey TD, Merlo DJ, Orr GL;  
 PI Petell J, Roberts JL, Rocheleau TA, Schoonover S;  
 PI Strickland JA;  
 XX  
 XX WPI; 1997-281022/25.  
 DR N-PSDB; AAT68836.  
 XX  
 PT Photorhabdus sp. insecticidal protein toxins and DNA encoding them -  
 PT can be genetically engineered into insect larvae food and plants for  
 PT insect control  
 XX  
 PS Claim 8; Page 119-129; 276pp; English.  
 XX  
 CC This polypeptide comprises a specifically claimed 280.6 kDa  
 CC insecticidal toxin, TcBA, of Photorhabdus luminescens that is  
 CC proteolytically cleaved into 207.6 kDa (see AAW18302) and 62.9 kDa  
 CC polypeptides (see AAW18303). The TcBA polypeptide can be expressed  
 CC in host cells using a gene (see AAT68836) isolated from a genomic  
 CC library. Claimed toxin proteins of P. luminescens (see AAW17871,  
 CC AAW17884-89, AAW17899-900, AAW18301-06) can be applied to, or  
 CC genetically engineered into, insect larvae food and plants for insect  
 CC control. The Photorhabdus toxins are particularly effective against  
 CC Southern corn rootworm, Colorado potato beetle, Western corn rootworm,  
 CC meal worm, boll weevil and turf grub (Coleoptera), beet armyworm, black  
 CC cutworm, cabbage looper, codling moth, corn earworm, European corn  
 CC borer, tobacco hornworm and tobacco budworm (Lepidoptera), and  
 CC are also active against insects of the orders Hymenoptera, Diptera  
 CC Dictyoptera, Acarina and Homoptera. (All claimed).  
 XX  
 SQ Sequence 2504 AA;  
 Alignment Scores:  
 Pred. No.: 2.5e-09 Length: 2504  
 Score: 126.50 Matches: 28  
 Percent Similarity: 64.71% Conservative: 5  
 Best Local Similarity: 54.90% Mismatches: 17  
 Query Match: 47.74% Indels: 1  
 DB: 18 Gaps: 1  
 US-09-856-221-2 (1-154) x AAW17871 (1-2504)  
 QY 1 CAACACCTTCGTGATCATTATACGATAAAATCGCGCACTCGCTCAACGGCGGCGCTTG 60  
 DB 1911 GlnGlnValArgGlnGlnValLeuThrGlnLeuArg---LeuAsnSerArgValLysThr 1929  
 QY 61 CGCGCAACCTACTCGTAATTCGTTAAACCGCTTTGTTCTCTCCTCAGATAAACAACAAA 120  
 DB 1930 ProLeuLeuGlyThrAlaAsnSerLeuThrAlaLeuPheLeuProGlnGlnAsnSerLys 1949  
 QY 121 CTGCAAACTTACTGCGACAGCTTAGCACACGC 153  
 DB 1950 LeuLysGlyTyrTrpArgThrLeuAlaGlnArg 1960  
 RESULT 3  
 AAW56557  
 ID AAW56557 standard; Protein; 2504 AA.  
 XX  
 AC AAW56557;  
 XX  
 XX 07-AUG-1998 (first entry)  
 XX  
 XX Toxin TcBA, encoded by the tcba gene from genomic region tcb.  
 DE  
 XX Photorhabdus luminescens W-14; nematode; symbiotic;  
 KW Heterorhabditis; tca; tcb; tcc; insecticidal activity; toxin;



KW Lepidoptera; Coleoptera; Hymenoptera; Diptera; Dictyoptera; Acarina;  
KW Homoptera; Southern; Western corn rootworm; Colorado potato beetle;  
KW mealworm; boll weevil; turf grub; beetle armyworm; black cutworm;  
KW cabbage looper; codling moth; corn earworm; European corn borer;  
KW Tobacco hornworm; budworm.  
XX  
OS Photorhabdus luminescens.  
XX  
PN W09808932-AL.  
XX  
PD 05-MAR-1998.  
XX  
XX  
PF 05-MAY-1997; 97WO-US07657.  
XX  
XX 06-NOV-1996; 96WO-US18003.  
PR 28-AUG-1996; 96US-0705484.  
PR 06-NOV-1996; 96US-0743699.  
XX  
XX (DOWC ) DOWELANCO.  
PA (WISC ) WISCONSIN ALUMNI RES FOUND.  
XX  
XX Blackburn MB, Bowen DJ, Ciche TA, Ensign JC, Fatig R;  
PI French-Constant RH, Guo L, Hey TD, Merlo DJ, Orr GL;  
PI Petell J, Roberts JL, Rocheleau TA, Schoonover S;  
PI Strickland JA, Sukhapinda K;  
XX  
XX WPI; 1998-179427/16.  
DR N-PSDB; AAV2985.  
XX  
XX Isolated toxins from Photorhabdus luminescens strains - useful for  
PT control of insect pests  
PT  
XX  
XX Claim 34; Pages 163-169; 321pp; English.  
XX  
CC The present sequence represents a protein named TcBa of the bacterium  
CC Photorhabdus luminescens (W-14). This is a symbiotic bacterium of the  
CC nematodes of the Heterorhabditis genus. The bacterium has at least 4  
CC distinct genomic regions, tca, tcb, tcc, and tcd. Peptide products are  
CC produced from these regions that are associated with insecticidal  
CC activity. The native toxins are secreted proteins. The proteins are  
CC toxic to insects upon exposure and especially when ingested. The  
CC nucleic acid sequence can be used to produce transgenic plants,  
CC baculoviruses or microbial hosts for toxin production. They can be used  
CC to control insects pests from the Lepidoptera, Coleoptera, Hymenoptera,  
CC Diptera, Dictyoptera, Acarina or Homoptera orders, especially the  
CC Southern or Western corn rootworm, Colorado potato beetle, mealworm,  
CC boll weevil, turf grub, beetle armyworm, black cutworm, cabbage looper,  
CC codling moth, corn earworm, European corn borer or tobacco hornworm  
CC or budworm.  
XX  
XX Sequence 2504 AA;  
SQ  
Alignment Scores:  
Pred. No.: 2.5e-09 Length: 2504  
Score: 126.50 Matches: 28  
Percent Similarity: 64.71% Conservative: 5  
Best Local Similarity: 54.90% Mismatches: 17  
Query Match: 47.74% Indels: 1  
DB: 19 Gaps: 1  
US-09-856-221-2 (1-154) x AAW56557 (1-2504)  
QY 1 CAACACATTCGTGATCATATTATCAGCATATAAATCGCGCACTCGCTCAACGGCGGCGCTTG 60  
Db 1911 GlnGlnValArgGlnGlnValLeuThrGlnLeuArg---LeuAsnSerArgValLysThr 1929  
QY 61 CCGCGCAACACGTACTGCAAAATTCGTTAAACCGCTTTGCTCTCAGATAAACAACAAAA 120  
Db 1930 ProLeuLeuGlyThrAlaAsnSerLeuThrAlaLeuPheLeuProGlnGluAsnSerLys 1949  
QY 121 CTCGAAGTTACTGGCAGCGTTAGCACAACGC 153  
Db 1950 LeuLysGlyTyrTTPArgThrLeuAlaGlnArg 1960

RESULT 4  
AAB72610  
ID AAB72610 standard; Protein; 2504 AA.  
XX  
AC AAB72610;  
XX  
DT 04-MAY-2001 (first entry)  
XX  
DE Photorhabdus tcbA toxin.  
XX  
KW TcBa; TcBa; insect toxin; plant; insect resistance.  
XX  
OS Photorhabdus sp.  
XX  
PN W0200111029-AL.  
XX  
PD 15-FEB-2001.  
XX  
PF 11-AUG-2000; 2000WO-US22237.  
XX  
PR 11-AUG-1999; 99US-0148356.  
XX  
PA (DOWC ) DOW AGROSCIENCES LLC.  
XX  
XX Petell JK, Merlo DJ, Herman RA, Roberts JL, Guo L, Schafer BW;  
PI Sukhapinda K, Merlo AO;  
XX  
XX WPI; 2001-191536/19.  
DR N-PSDB; AAF58779.  
XX  
XX Novel polynucleotide sequence encoding insect toxins, useful for  
PT producing transgenic plants having resistance to insects, especially  
PT corn rootworm  
PT  
XX  
XX Disclosure; Page 62-72; 106pp; English.  
XX  
CC The present invention provides the protein and coding sequences of  
CC modified versions of the Photorhabdus TcBa and TcBa toxins. These are  
CC suitable for expression in plants. The toxins are effective against  
CC insects upon ingestion, and the sequences provided can be used to produce  
CC transgenic plants with insect resistance. The present sequence is the  
CC Photorhabdus TcBa protein.  
XX  
XX Sequence 2504 AA;  
SQ  
Alignment Scores:  
Pred. No.: 2.5e-09 Length: 2504  
Score: 126.50 Matches: 28  
Percent Similarity: 64.71% Conservative: 5  
Best Local Similarity: 54.90% Mismatches: 17  
Query Match: 47.74% Indels: 1  
DB: 22 Gaps: 1  
US-09-856-221-2 (1-154) x AAB72610 (1-2504)  
QY 1 CAACACATTCGTGATCATATTATCAGCATATAAATCGCGCACTCGCTCAACGGCGGCGCTTG 60  
Db 1911 GlnGlnValArgGlnGlnValLeuThrGlnLeuArg---LeuAsnSerArgValLysThr 1929  
QY 61 CCGCGCAACACGTACTGCAAAATTCGTTAAACCGCTTTGCTCTCAGATAAACAACAAAA 120  
Db 1930 ProLeuLeuGlyThrAlaAsnSerLeuThrAlaLeuPheLeuProGlnGluAsnSerLys 1949  
QY 121 CTCGAAGTTACTGGCAGCGTTAGCACAACGC 153  
Db 1950 LeuLysGlyTyrTTPArgThrLeuAlaGlnArg 1960  
RESULT 5  
AAB72612  
ID AAB72612 standard; Protein; 2505 AA.  
XX  
AC AAB72612;

XX 04-MAY-2001 (first entry)  
XX Modified Photorhabdus tcbA toxin.  
XX TcdA; TcbA; insect toxin; plant; insect resistance.  
XX Photorhabdus sp.  
XX Synthetic.  
XX WO200111029-A1.  
XX 15-FEB-2001.  
XX 11-AUG-2000; 2000WO-US22237.  
XX 11-AUG-1999; 99US-0148356.  
XX (DOWC ) DOW AGROSCIENCES LLC.  
XX Petell JK, Merlo DJ, Herman RA, Roberts JL, Guo L, Schafer BW;  
XX Sukhapinda K, Merlo AO;  
XX WPI; 2001-191536/19.  
XX N-PSDB; AAF58781.  
XX Novel polynucleotide sequence encoding insect toxins, useful for  
XX producing transgenic plants having resistance to insects, especially  
XX corn rootworm  
XX Claim 1; Page 83-93; 106pp; English.  
XX The present invention provides the protein and coding sequences of  
XX modified versions of the Photorhabdus TcdA and tcbA toxins. These are  
XX suitable for expression in plants. The toxins are effective against  
XX insects upon ingestion, and the sequences provided can be used to produce  
XX transgenic plants with insect resistance. The present sequence is the  
XX modified Photorhabdus TcdA protein.  
XX Sequence 2505 AA;  
Alignment Scores:  
Pred. No.: 2.5e-09 Length: 2505  
Score: 126.50 Matches: 28  
Percent Similarity: 64.71% Conservative: 5  
Best Local Similarity: 54.90% Mismatches: 17  
Query Match: 47.74% Indels: 1  
DB: 22 Gaps: 1  
US-09-856-221-2 (1-154) x AAB72612 (1-2505)  
QY 1 CAACCATTCGFGATCATATACGATAAAATCGGCAACTGCGTCAACGCGCGCTTG 60  
Db 1912 GlnGlnValArgGlnGlnValLeuThrGlnLeuArg--LeuAsnSerArgValLysThr 1930  
QY 61 CGCGGGAACGTAAGTAAATCGTTAAACCGCTTTGTTCTTCCTCAGATAACAAAAA 120  
Db 1931 ProLeuLeuGlyThrAlaAsnSerLeuThrAlaLeuPheLeuProGlnGlnAsnSerLys 1950  
QY 121 CTCGAAGTTACTGCGACAGCTAGCACACGC 153  
Db 1951 LeuLysGlyTyrIrrPArgThrLeuAlaGlnArg 1961  
RESULT 6  
AAW17899  
ID AAW17899 standard; Protein: 2516 AA.  
XX  
XX AAW17899;  
XX  
XX 29-JAN-1998 (first entry)  
XX Photorhabdus luminescens insect toxin TcdA.  
XX

KW Insecticide; insect; toxin; pest control; biological control;  
KW Photorhabdus luminescens; TcdA; Southern corn rootworm;  
KW Colorado potato beetle; Western corn rootworm; meal worm;  
KW boll weevil; turf grub; Coleoptera; beet armyworm; black cutworm;  
KW cabbage looper; codling moth; corn earworm; European corn borer;  
KW tobacco hornworm; tobacco budworm; Lepidoptera; Hymenoptera;  
KW Diptera, Dictyoptera; Acarina; Homoptera.  
XX  
XX Photorhabdus luminescens strain W-14 (ATCC 55397).  
XX  
XX Key Location/Qualifiers  
FH Protein 1..2516  
FT /label= TcdA  
FT Protein 89..1937  
FT /label= TcdAii  
FT Peptide 89..100  
FT /note= "S2 N-terminus (Claim 30)"  
FT Peptide 284..299  
FT /note= "tryptic peptide (Claim 30)"  
FT Peptide 554..563  
FT /note= "tryptic peptide (Claim 30)"  
FT Peptide 1080..1092  
FT /note= "isolated N-terminal peptide (Claim 30)"  
FT Peptide 1385..1400  
FT /note= "tryptic peptide (Claim 30)"  
FT Peptide 1478..1497  
FT /note= "tryptic peptide (Claim 30)"  
FT Peptide 1620..1642  
FT /note= "claimed peptide (Claim 30)"  
FT Peptide 1938..1948  
FT /note= "N-terminal peptide (Claim 30)"  
FT Protein 1938..2516  
FT /label= TcdAiii  
FT Peptide 2327..2345  
FT /note= "tryptic peptide (Claim 30)"  
FT Peptide 2398..2408  
FT /note= "tryptic peptide (Claim 30)"  
XX  
XX WO9717432-A1.  
XX  
XX 15-MAY-1997.  
XX  
XX 06-NOV-1996; 96WO-US18003.  
XX  
XX 28-AUG-1996; 96US-0705484.  
XX 06-NOV-1995; 95US-0007255.  
XX 28-FEB-1996; 96US-0608423.  
XX  
XX (WISC ) WISCONSIN ALUMNI RES FOUND.  
XX  
XX Blackburn MB, Bowen DJ, Cliche TA, Ensign JC, Fatig R;  
XX French-Constant RH, Guo L, Hey TD, Merlo DJ, Orr GL;  
XX Petell J, Roberts JL, Rocheleau TA, Schoonover S;  
XX Strickland JA;  
XX  
XX WPI: 1997-281022/25.  
XX N-PSDB; AAT68843.  
XX  
XX Photorhabdus sp. insecticidal protein toxins and DNA encoding them -  
XX can be genetically engineered into insect larvae food and plants for  
XX insect control  
XX  
XX Claim 34; Page 185-194; 276pp; English.  
XX  
XX This polypeptide comprises the 282.9 kDa TcdA insecticidal toxin  
XX protein of Photorhabdus luminescens W-14. Its sequence was deduced  
XX from a genomic DNA clone (AAW68843) and includes N-terminal and  
XX tryptic peptide sequences obtained from the isolated protein. The  
XX 282.9 kDa insect toxin is proteolytically processed into TcdAii and  
XX TcdAiii components (see AAW17900 and AAW18301). Claimed toxins of P.  
XX luminescens (see AAW17871, AAW17884-89, AAW17899-900, AAW18301-06) can  
XX be produced by recombinant DNA methods and applied to, or genetically  
XX engineered into, insect larvae food and plants for insect control.  
XX

AA	Sequence	2516 AA;
SQ		

CC activity. The native toxins are secreted proteins. The proteins are  
CC toxic to insects upon exposure and especially when ingested. The  
CC nucleic acid sequence can be used to produce transgenic plants,  
CC baculoviruses or microbial hosts for toxin production. They can  
CC be used to control insects pests from the Lepidoptera, Coleoptera, Hymenoptera  
CC (Diptera, Dictyoptera, Acarina or Homoptera orders, especially the  
CC Southern or Western corn rootworm, Colorado potato beetle, mealworm,  
CC boll weevil, turf grub, beetle armyworm, black cutworm, cabbage looper,  
CC codling moth, corn earworm, European corn borer or tobacco hornworm  
CC or budworm.

Db 1914 ThrThrGlnAsnAlaHisAspSerAlaIleValAlaLeuArgGlnAsnIleProThrPro 1933

QY 64 GCGAAA-----CGTACTGCAAATTCGTTAACCGCCTTGTTCCTTCCTCAGATAAAC 114

Db 1934 AlaProLeuSerLeuAlaGserAlaAsnThrLeuThrAspLeuPheLeuProGlnLeuAsn 1953

**THE UNIVERSITY OF CHICAGO**

Db 1954 GluValMetMetAsnTyrTrpGlnThrLeuAlaGlnArg 1966

AAW56572

XX  
-----XX  
DE  
07-AUG-1969 (4:54 PM)  
XX

DEFXX Toxin TcdA: encoded by the tcdA gene from genomic region tcd

XX  
KW  
Photorhabdus luminescens W-14: nematode: symbiotic.

KW Lepidoptera; Coleoptera; Hymenoptera; Diptera; Dictyoptera; Acarina;  
KW Homoptera; Southern; Western corn rootworm; Colorado potato beetle;  
KW Homoptera; Southern; Western corn rootworm; Colorado potato beetle;

KW cabbage looper; codling moth; corn earworm; European corn borer;  
Tobacco hornworm. budworm

OS Photorhabdus luminescens.

XX  
XX  
XX

PF 05-MAY-1997; 97WO-US07657.

FR 00 NOV 1990; 90WC-US18003;  
PR 28-AUG-1996; 96US-0705484.

AA  
PA (DOWC ) DOWELANCO.

XX  
PI  
Blackburn MB. Bowen D.J. Cliche TA. Ensian JC. Fatic R.

PI Petell J, Roberts JL, Rocheleau TA, Schoonover S;  
PI Strickland JA, Sukhaninda K.

DR WPI; 1998-179427/16.  
DR N-PCNB. AAU20028

DR N-PSDB; AAF58778.

XX Novel polynucleotide sequence encoding insect toxins, useful for

PT producing transgenic plants having resistance to insects, especially

PT corn rootworm

XX Disclosure; Page 51-61; 106pp; English.

XX The present invention provides the protein and coding sequences of

CC modified versions of the photorhabdus TcdA and TcBA toxins. These are

CC suitable for expression in plants. The toxins are effective against

CC insects upon ingestion, and the sequences provided can be used to produce

CC transgenic plants with insect resistance. The present sequence is the

CC photorhabdus TcdA protein.

XX Sequence 2516 AA;

Alignment Scores:

Pred. No.: 3,69e-08 Length: 2516

Score: 118.50 Matches: 26

Percent Similarity: 66.04% Conservatives: 9

Best Local Similarity: 49.06% Mismatches: 15

Query Match: 44.72% Indels: 3

DB: 22 Gaps: 1

US-09-856-221-2 (1-154) x AAB72609 (1-2516)

QY 4 ACCATTCGTGATTCATATACGATAAAATCGCGCAACTGCGTCAACGCGCGCTTGCGG 63

Db 1914 ThrThrGlnAsnAlaHisAspSerAlaIleValAlaLeuArgGlnAsnIleProThrPro 1933

QY 64 GCGAAA-----CGTACTGCAAAATTCGTTAAACCGCTTTGTTCCCTCCTCAGATAAAC 114

Db 1934 AlaProLeuSerLeuArgSerAlaAsnThrLeuThrAspLeuPheLeuProGlnIleAsn 1953

QY 115 AAAAAGTCAAGTCTACTGCGAGCGTTAGCACACGC 153

Db 1954 GluValMetMetAsnTyrTrpGlnThrLeuAlaGlnArg 1966

RESULT 9

AAB72611

ID AAB72611 standard; Protein; 2517 AA.

XX

AC AAB72611;

XX

DT 04-MAY-2001 (first entry)

XX

DE Modified Photorhabdus tcdA toxin.

XX

KW TcdA; TcBA; insect toxin; plant; insect resistance.

XX

OS Photorhabdus sp.

XX

OS Synthetic.

XX

PN WO200111029-A1.

XX

PD 15-FEB-2001.

XX

PF 11-AUG-2000; 2000WO-US22237.

XX

PR 11-AUG-1999; 99US-0148356.

XX

PA (DOWC ) DOW AGROSCIENCES LLC.

XX

PI Petell JK, Merlo DJ, Herman RA, Roberts JL, Guo L, Schafer BW;

PI Sukhapinda K, Merlo AO;

XX

DR WPI; 2001-191536/19.

XX

DR N-PSDB; AAF58780.

XX

PT Novel polynucleotide sequence encoding insect toxins, useful for

PT producing transgenic plants having resistance to insects, especially

PT corn rootworm

XX Claim 1; Page 72-83; 106pp; English.

XX The present invention provides the protein and coding sequences of

CC modified versions of the photorhabdus TcdA and TcBA toxins. These are

CC suitable for expression in plants. The toxins are effective against

CC insects upon ingestion, and the sequences provided can be used to produce

CC transgenic plants with insect resistance. The present sequence is the

CC modified Photorhabdus TcdA protein.

XX Sequence 2517 AA;

Alignment Scores:

Pred. No.: 3,69e-08 Length: 2517

Score: 118.50 Matches: 26

Percent Similarity: 66.04% Conservatives: 9

Best Local Similarity: 49.06% Mismatches: 15

Query Match: 44.72% Indels: 3

DB: 22 Gaps: 1

US-09-856-221-2 (1-154) x AAB72611 (1-2517)

QY 4 ACCATTCGTGATTCATATACGATAAAATCGCGCAACTGCGTCAACGCGCGCTTGCGG 63

Db 1915 ThrThrGlnAsnAlaHisAspSerAlaIleValAlaLeuArgGlnAsnIleProThrPro 1934

QY 64 GCGAAA-----CGTACTGCAAAATTCGTTAAACCGCTTTGTTCCCTCCTCAGATAAAC 114

Db 1935 AlaProLeuSerLeuArgSerAlaAsnThrLeuThrAspLeuPheLeuProGlnIleAsn 1954

QY 115 AAAAAGTCAAGTCTACTGCGAGCGTTAGCACACGC 153

Db 1955 GluValMetMetAsnTyrTrpGlnThrLeuAlaGlnArg 1967

RESULT 10

AAB72614

ID AAB72614 standard; Protein; 2537 AA.

XX

AC AAB72614;

XX

DT 04-MAY-2001 (first entry)

XX

DE TcdA toxin-zeln ER signal peptide fusion protein.

XX

KW TcdA; TcBA; insect toxin; plant; insect resistance.

XX

OS Chimeric - Photorhabdus sp.

XX

OS Chimeric - Zea mays.

XX

PN WO200111029-A1.

XX

PD 15-FEB-2001.

XX

PF 11-AUG-2000; 2000WO-US22237.

XX

PR 11-AUG-1999; 99US-0148356.

XX

PA (DOWC ) DOW AGROSCIENCES LLC.

XX

PI Petell JK, Merlo DJ, Herman RA, Roberts JL, Guo L, Schafer BW;

PI Sukhapinda K, Merlo AO;

XX

DR WPI; 2001-191536/19.

XX

DR N-PSDB; AAF58783.

XX

PT Novel polynucleotide sequence encoding insect toxins, useful for

PT producing transgenic plants having resistance to insects, especially

PT corn rootworm

XX

PS Example 1; Page 94-104; 106pp; English.

XX The present invention provides the protein and coding sequences of

CC modified versions of the photorhabdus TcdA and TcBA toxins. These are

CC suitable for expression in plants. The toxins are effective against  
 CC insects upon ingestion, and the sequences provided can be used to produce  
 CC transgenic plants with insect resistance. The present sequence is a  
 CC fusion protein comprising the Photobabidus TcdA toxin and the maize zein  
 CC protein ER signal peptide.

XX SQ Sequence 2537 AA;

Alignment Scores:  
 Pred. No.: 3.69e-08 Length: 2537  
 Score: 118.50 Matches: 26  
 Percent Similarity: 66.04% Conservative: 9  
 Best Local Similarity: 49.06% Mismatches: 15  
 Query Match: 44.72% Indels: 3  
 DB: 22 Gaps: 1

US-09-856-221-2 (1-154) x AAB72614 (1-2537)

QY 4 ACCATTCGTGATCATATTCAGCATATAAATGCGCAACTGCGTCAACGCGCGCTTGCGG 63  
 Db 1935 ThrThrGlnAsnAlaHisAspSerAlaIleValAlaLeuArgGlnAsnIleProThrPro 1954  
 QY 64 GCGAAA-----CGTACTGCAAAATTCGTTACCGCTTTGTTCTCTCAGATAAAC 114  
 Db 1955 AlaProLeuSerLeuArgSerAlaAsnThrLeuThrAspLeuPheLeuProGlnIleAsn 1974  
 QY 115 AAAAACTGCAAGTTACTGCGACAGCTTAGCACACGC 153  
 Db 1975 GluValMetMetAsnTyrrpGlnThrLeuAlaGlnArg 1987

RESULT 11

AAW18303  
 ID AAW18303 standard; Protein: 573 AA.

XX AC AAW18303;  
 XX DT 09-NOV-1999 (first entry)  
 XX DE Photobabidus luminescens hph2-encoded insecticidal toxin.

XX KW Symbiotic bacterium; nematode; insect; larva; toxin; insecticide.

XX OS Photobabidus luminescens.

XX PN W09942589-A2.

XX PD 26-AUG-1999.

XX PF 18-FEB-1999; 99WO-EP01015.

XX PR 20-JAN-1999; 99US-0116439.

XX PR 20-FEB-1998; 98US-0027080.

XX PA (NOVS ) NOVARTIS AG.

XX PA (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.

XX PI Anderson AR, Chen JS, Dunn MM, Hart HP, Kramer VC;

XX PI Morgan MK, Warren GW;

XX DR WPI; 1999-527479/44.

XX DR N-PSDB; AA206831.

XX PT New nucleic acid from Photobabidus luminescens encoding insecticidal  
 XX toxins, used for making resistant transgenic plants

XX PS Claim 26; Page 133-140; 148pp; English.

XX CC This sequence represents the hph2 gene encoded insecticidal  
 CC toxin from Photobabidus luminescens. It is one of three insecticidal  
 CC toxins (AAV33728-Y33730) encoded by open reading frames (orfs) in a 38kb  
 CC fragment of P. luminescens DNA (AA206831). The hph2 gene was identified  
 CC using probe #2 which was amplified from P. luminescens genomic DNA using  
 CC PCR primers AA206829-206830. P. luminescens is a member of the

CC Enterobacteriaceae family and is a symbiotic bacterium of  
 CC nematodes of the genus Heterorhabdus. The nematodes colonise  
 CC insect larvae, kill them, and their offspring feed on the dead  
 CC larvae. However, the insecticidal agents are produced by P.  
 CC luminescens rather than the nematodes. The toxins have activity against  
 CC Lepidopteran insects such as Cabbage Looper (Trichoplusia ni),  
 CC European Corn Borer (Ostrinia nubilalis) and Fall Armyworm  
 CC (Spodoptera frugiperda) and also against Coleopteran insects  
 CC (e.g., Colorado Potato Beetle, Leptinotarsa decimlineata). In  
 CC addition the toxins are active against strains resistant to known  
 CC insecticides. The DNA sequence can be used to generate transgenic plants  
 CC of various species that are resistant to economically important insect  
 CC pests and also for recombinant production of the toxins for use as  
 CC insecticides.

XX SQ Sequence 2522 AA;

Alignment Scores:

Pred. No.: 1.01e-07 Length: 2522  
 Score: 115.50 Matches: 26  
 Percent Similarity: 68.63% Conservative: 9  
 Best Local Similarity: 50.98% Mismatches: 15  
 Query Match: 43.58% Indels: 1  
 DB: 20 Gaps: 1

US-09-856-221-2 (1-154) x AAY33729 (1-2522)

QY 4 ACCATTCGTGATCATATTCAGCATATAAATGCGCAACTGCGTCAACGCGGCGG 60

Db 1922 ThrThrGlnSerAlaHisSerSerIleValAlaLeuArgGlnSerThrProAlaLeu 1941

QY 61 CCGCGCAACGTACTGCAAAATTCGTTAACCGCTTTGTTCTCTCAGATAACAAAAA 120

Db 1942 LeuSerLeuArgSerAlaAsnThrLeuThrAspLeuPheLeuProGlnIleAsnGluVal 1961

QY 121 CTGCNAGTTACTGCGACAGCTTAGCACACGC 153

Db 1962 MetMetAsnTyrrpGlnThrLeuAlaGlnArg 1972

RESULT 12

AAW18303  
 ID AAW18303 standard; Protein: 573 AA.

XX AC AAW18303;

XX DT 30-JAN-1998 (first entry)

XX DE Photobabidus luminescens insect toxin TcbAiii.

XX KW Insecticide; insect; toxin; pest control; biological control;

XX KW Photobabidus luminescens; TcbA; Southern corn rootworm;

XX KW Colorado potato beetle; Western corn rootworm; meal worm;

XX KW boll weevil; turf grub; Coleoptera; beet armyworm; black cutworm;

XX KW cabbage looper; codling moth; corn earworm; European corn borer;

XX KW tobacco hornworm; tobacco budworm; Lepidoptera; Hymenoptera;

XX KW Diptera, Dictyoptera; Acarina; Homoptera.

XX OS Photobabidus luminescens strain W-14 (ATCC 55397).

XX PN W09717432-A1.

XX PD 15-MAY-1997.

XX PF 06-NOV-1996; 96WO-US18003.

XX PR 28-AUG-1996; 96US-0705484.

XX PR 06-NOV-1995; 95US-0007255.

XX PR 28-FEB-1996; 96US-0608423.

XX PA (WISC ) WISCONSIN ALUMNI RES FOUND.

XX PI Blackburn MB, Bowen DJ, Cliche TA, Ensign JC, Fatig R;

XX PI French-Constant RH, Guo L, Hey TD, Merlo DJ, Orr GL;





Thu Jan 16 09:33:21 2003

us-09-856-221-2.rag

Page 10

Db 22 TyrTrpGlnThrLeuAlaGlnArg 29

Search completed: January 15, 2003, 15:50:28  
Job time : 19.2569 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: January 15, 2003, 15:45:23 : Search time 4.96166 Seconds  
(without alignments)  
1826.456 Million cell updates/sec

Title: US-09-856-221-2  
Perfect score: 265  
Sequence: 1 caaacattgtgatcatta.....gcagacgttagcacacgcc 154

Scoring table: BIOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 262574 segs, 29422922 residues

Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+n2p.model  
-Q=/cgn2\_1/USPTO\_spool/US09856221/runat\_15012003\_153926\_12521/app.query.fasta\_1.1948  
-DB=Issued\_Patents\_AA -OPMT=fastan -SUFFIX=rai -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdl  
-LIST=45 -DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09856221\_cgn\_1\_24\_runat\_15012003\_153926\_12521 -NCPU=6 -ICPU=3  
-NO\_XLPY -NO\_MMAP -LARGQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THRADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued\_Patents\_AA.\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	115.5	43.6	2522	4	US-09-251-645-13
c 2	56.5	21.2	951	1	Sequence 13, Appl
3	56	21.1	355	4	Sequence 2, Appl1
c 4	55	20.6	771	4	Sequence 14, Appl
c 5	53	19.9	329	4	Sequence 2, Appl1
6	52.5	19.8	443	1	Sequence 19, Appl
7	52.5	19.8	443	2	Sequence 2, Appl1
8	52.5	19.8	1048	4	Sequence 2, Appl1
c 9	51.5	19.3	227	3	Sequence 85, Appl
10	51	19.2	236	4	Sequence 45, Appl
11	51	19.2	236	4	Sequence 4, Appl1
12	51	19.2	236	4	Sequence 2, Appl1

13	51	19.2	616	4	US-09-134-001C-4618	Sequence 4618, Ap
14	50.5	19.1	575	4	US-09-171-461-7	Sequence 7, Appl1
c 15	50.5	18.9	984	2	US-08-673-789-6	Sequence 6, Appl1
c 16	50.5	18.9	1457	3	US-08-665-259-27	Sequence 27, Appl
c 17	50.5	18.9	1457	3	US-08-762-500-27	Sequence 27, Appl
c 18	50	18.7	33	4	US-08-981-392-63	Sequence 63, Appl
c 19	50	18.7	66	3	US-08-908-643C-11	Sequence 11, Appl
c 20	50	18.7	76	3	US-08-908-643C-9	Sequence 9, Appl1
c 21	50	18.7	92	3	US-08-908-643C-7	Sequence 7, Appl1
c 22	50	18.7	124	3	US-08-908-643C-5	Sequence 5, Appl1
c 23	50	18.7	257	3	US-08-637-759B-92	Sequence 92, Appl
c 24	50	18.7	257	3	US-08-871-355A-92	Sequence 92, Appl
c 25	50	18.7	257	4	US-09-201-945-92	Sequence 92, Appl
c 26	50	18.7	494	4	US-08-378-313-23	Sequence 23, Appl
c 27	50	18.7	494	4	US-08-378-313-23	Sequence 23, Appl
28	50	18.9	698	4	US-09-134-001C-3632	Sequence 3632, Ap
29	49.5	18.7	148	1	US-08-468-347-22	Sequence 22, Appl
30	49.5	18.7	148	2	US-08-467-389-22	Sequence 22, Appl
31	49.5	18.7	148	2	US-08-779-379-22	Sequence 22, Appl
32	49.5	18.7	148	2	US-08-469-219-22	Sequence 22, Appl
33	49.5	18.7	148	4	US-09-228-152-22	Sequence 22, Appl
c 34	49.5	18.5	339	3	US-08-758-280-1	Sequence 1, Appl1
c 35	49.5	18.5	339	3	US-08-758-280-2	Sequence 2, Appl1
c 36	49.5	18.5	339	3	US-08-964-614A-1	Sequence 1, Appl1
c 37	49.5	18.5	339	3	US-08-964-614A-2	Sequence 2, Appl1
38	49.5	18.7	374	1	US-08-468-847B-12	Sequence 12, Appl
39	49.5	18.7	375	2	US-08-459-101A-2	Sequence 2, Appl1
c 40	49.5	18.5	3165	2	US-08-459-146-3	Sequence 3, Appl1
c 41	49.5	18.5	3165	2	US-08-459-065-3	Sequence 3, Appl1
42	49	18.5	329	4	US-09-011-769A-51	Sequence 51, Appl
43	49	18.5	349	4	US-09-011-769A-47	Sequence 47, Appl
44	49	18.5	349	4	US-09-011-769A-60	Sequence 60, Appl
45	49	18.5	349	4	US-09-011-769A-64	Sequence 64, Appl

ALIGNMENTS

RESULT 1  
US-09-251-645-13  
; Sequence 13, Application US/09251645  
; Patent No. 6281413  
; GENERAL INFORMATION:  
; APPLICANT: Kramer, Vance C.  
; APPLICANT: Morgan, Michael K.  
; APPLICANT: Anderson, Arne R.  
; APPLICANT: Hart, Hope  
; APPLICANT: Warren, Gregory W.  
; APPLICANT: Dunn, Martha  
; APPLICANT: Chen, Jeng S.  
; TITLE OF INVENTION: NOVEL INSECTICIDAL TOXINS FROM PHOTORHABDUS LUMINESCENS  
; FILE REFERENCE: CGC1963/A  
; CURRENT APPLICATION NUMBER: US/09/251,645  
; CURRENT FILING DATE: 1999-02-17  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 13  
; LENGTH: 2522  
; TYPE: PRT  
; ORGANISM: Photorhabdus luminescens  
US-09-251-645-13

Alignment Scores:	1.19e-08	Length:	2522
Pred. No.:	115.50	Matches:	26
Score:	68.63%	Conservative:	9
Best Local Similarity:	50.98%	Mismatches:	15
Query Match:	43.58%	Indels:	1
DB:	4	Gaps:	1

US-09-856-221-2 (1-154) x US-09-251-645-13 (1-2522)

QY 4 ACCATTGCTGATCATTCAGCATATAATGCGCAACTGCGTCAACGCGG---GCCTTG 60

```
Db 1922 ThrGlnSerAlaHisSerSerIleValAlaLeuArgGlnSerThrProAlaLeu 1941
      |||   |||   |||   |||   |||   |||   |||   |||   |||   |||
QY 61 CCGCGAAGCTACTGCAATTCGTTAACCCTGTTGTCCTCTCAGATAACAAAAA 120
      ::|   |||   |||   |||   |||   |||   |||   |||   |||   |||
Db 1942 LeuSerLeuArgSerAlaAsnThrLeuThrAspLeuPheLeuProGlnIleAsnGluVal 1961
      ::|   |||   |||   |||   |||   |||   |||   |||   |||   |||
QY 121 CTGCAAGTCTACTGGCAGACGTTAGCACAACGC 153
      ::|   |||   |||   |||   |||   |||   |||   |||   |||   |||
Db 1962 MetMetAsnTyrTrpGlnThrLeuAlaGlnArg 1972
      ::|   |||   |||   |||   |||   |||   |||   |||   |||   |||
RESULT 2
US-08-162-809-2
; Sequence 2, Application US/08162809
; Patent No. 5457048
; GENERAL INFORMATION:
; APPLICANT: Pasquale, Elena B.
; TITLE OF INVENTION: Sajjadi, Fereydoun G.
; TITLE OF INVENTION: NOVEL EPH-RELATED TYROSINE KINASES,
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES, AND METHODS OF USE
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CAMPBELL AND FLORES
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States of America
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/162,809
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 9503
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 951 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-162-809-2
Alignment Scores:
Pred. No.: 6.39 Length: 951
Score: 56.50 Matches: 17
Percent Similarity: 41.51% Conservative: 5
Best Local Similarity: 32.08% Mismatches: 27
Query Match: 21.16% Indels: 4
DB: 1 Gaps: 1
US-09-856-221-2 (1-154) x US-08-162-809-2 (1-951)
QY 152 CGTTGCTGCTACCTGCTCCAGTACTTTCAGCTTTT-----TGTATTCTGAGGA 102
      |||   |||   |||   |||   |||   |||   |||   |||   |||   |||
Db 143 lysCysProSerValValGln-AsnPheAlaIlePheProGluThrMetThrGlyAlaG1 162
      |||   |||   |||   |||   |||   |||   |||   |||   |||   |||
QY 101 AGGAACAAAGCGTTACGAATTTGACGTACGTTTCGCGCGCAAGCGCGCGTTGACGC 42
      |||   |||   |||   |||   |||   |||   |||   |||   |||   |||
Db 162 userThrSerLeuValThrAlaArgGlyThrCysIleProAsnAlaGluGluValAspVa 182
      |||   |||   |||   |||   |||   |||   |||   |||   |||   |||
QY 41 AGTTCGCCCATTTTATGCTGATATGATCACCAGTGG 5
      |||   |||   |||   |||   |||   |||   |||   |||   |||   |||
Db 182 lProIleLysLeuTyrCysAsnGlyAspGlyGluTrp 194
      |||   |||   |||   |||   |||   |||   |||   |||   |||   |||
```

```
RESULT 3
US-08-984-618-14
; Sequence 14, Application US/08984618
; Patent No. 6251647
; GENERAL INFORMATION:
; APPLICANT: de Lencastre, Hermunia
; APPLICANT: Tomasz, Alexander
; TITLE OF INVENTION: AUXILIARY GENES AND PROTEINS OF
; TITLE OF INVENTION: METHICILLIN RESISTANT BACTERIA AND ANTAGONISTS THEREOF
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/984,618
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-213 CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 355 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: E. coli
US-08-984-618-14
Alignment Scores:
Pred. No.: 5.61 Length: 355
Score: 56.00 Matches: 18
Percent Similarity: 37.14% Conservative: 8
Best Local Similarity: 25.71% Mismatches: 22
Query Match: 21.13% Indels: 22
DB: 4 Gaps: 2
US-09-856-221-2 (1-154) x US-08-984-618-14 (1-355)
QY 1 CAACCATTCGTGATCATTTAT-----CAGCATAAATGCGCAACTG 42
      |||   |||   |||   |||   |||   |||   |||   |||   |||   |||
Db 224 GlnSerValGluGlnAlaTyrAlaGluAlaGlyGlnProGlnHisLysValThrGluPhe 243
      |||   |||   |||   |||   |||   |||   |||   |||   |||   |||
QY 43 -----CGTCAACGCGCG 54
      |||   |||   |||   |||   |||   |||   |||   |||   |||   |||
Db 244 IleAspAspMetAlaAlaAlaTyrAlaTrpAlaAspValValCysArgSerGlyAla 263
      |||   |||   |||   |||   |||   |||   |||   |||   |||   |||
QY 55 GCCTTGCGCGGCAAGCTACTGCAAAATTCGTTAACCGCTTTGTTCTCTCAGATAAAC 114
      ::|   |||   |||   |||   |||   |||   |||   |||   |||   |||
Db 264 LeuThrValSerGluIleAlaAlaGlyLeuProAlaLeuPheValPropheGlnHis 283
      ::|   |||   |||   |||   |||   |||   |||   |||   |||   |||
QY 115 AAAAAGTGCAGAAAGTTACTGCGACGCTTA 144
      |||   |||   |||   |||   |||   |||   |||   |||   |||   |||
Db 284 LysAspArgGlnGlnTyrTrpAsnAlaLeu 293
      |||   |||   |||   |||   |||   |||   |||   |||   |||   |||
```

## RESULT 4

US-09-462-284-2  
 ; Sequence 2, Application US/09462284  
 ; Patent No. 6309868  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Nestec S.A.  
 ; APPLICANT: Monod, Michel  
 ; APPLICANT: Doumas, Agnes  
 ; APPLICANT: Affolter, Micheal  
 ; APPLICANT: Van Den Broek, Peter  
 ; TITLE OF INVENTION: CLONING OF THE  
 ; TITLE OF INVENTION: PROLYL-DIPEPTIDYL-PEPTIDASE FROM  
 ; TITLE OF INVENTION: ASPERGILLUS ORYZAE  
 ; FILE REFERENCE: 8265-298  
 ; CURRENT APPLICATION NUMBER: US/09/462,284  
 ; CURRENT FILING DATE: 2000-01-03  
 ; NUMBER OF SEQ ID NOS: 9  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 2  
 ; LENGTH: 771  
 ; TYPE: PRT  
 ; ORGANISM: Fungus  
 US-09-462-284-2

Alignment Scores:  
 Pred. No.: 10.1 Length: 771  
 Score: 55.00 Matches: 17  
 Percent Similarity: 49.09% Conservative: 10  
 Best Local Similarity: 30.91% Mismatches: 20  
 Query Match: 20.60% Indels: 8  
 DB: 4 Gaps: 4

US-09-856-221-2 (1-154) x US-09-462-284-2 (1-771)

QY 151 GTTGTGCTACGCTGCCAGTACTTTCAGCTTTTGTATCTGAGGAGGACAA--- 95  
 DB 317 ValAlaValAspThrAlaSerAsnLysAlaThrValIleSerAspArgAspGlyThrAsp 336  
 QY 94 -----AAGCGGTAAACGAATTGTCAGTACGTTTCGCCGCAAGCGCGCGTGGAC 44  
 DB 337 GlyTrpLeuAspAsnLeuLeuSerMetLysTrpIleGlyProIleLysPro-----Ser 354  
 QY 43 GCAGTTCGCCCATTTATCTCAT---AATGATCAGCA---TGG 5  
 DB 355 AspLysAspAlaTrpTrpIleAspIleSerAspHisSerGlyTrp 369

## RESULT 5

US-09-651-200-19  
 ; Sequence 19, Application US/09651200  
 ; Patent No. 6429303  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Green et al  
 ; TITLE OF INVENTION: Polynucleotides Encoding Members of the Human B  
 ; TITLE OF INVENTION: Lymphocyte Activation Antigen B-7 Family and  
 ; TITLE OF INVENTION: Polypeptides Encoded Thereby  
 ; FILE REFERENCE: 15966-562 (CURA-62)  
 ; CURRENT APPLICATION NUMBER: US/09/651,200  
 ; CURRENT FILING DATE: 2000-08-30  
 ; PRIOR APPLICATION NUMBER: 60/152383  
 ; PRIOR FILING DATE: 1999-09-03  
 ; PRIOR APPLICATION NUMBER: 60/172909  
 ; PRIOR FILING DATE: 1999-12-21  
 ; PRIOR APPLICATION NUMBER: 60/183578  
 ; PRIOR FILING DATE: 2000-02-18  
 ; NUMBER OF SEQ ID NOS: 25  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 19  
 ; LENGTH: 329  
 ; TYPE: PRT  
 ; ORGANISM: Canis familiaris  
 US-09-651-200-19

Alignment Scores:

Pred. No.: 15.5 Length: 329  
 Score: 53.00 Matches: 13  
 Percent Similarity: 50.00% Conservative: 8  
 Best Local Similarity: 30.95% Mismatches: 15  
 Query Match: 19.85% Indels: 6  
 DB: 4 Gaps: 1

US-09-856-221-2 (1-154) x US-09-651-200-19 (1-329)

QY 136 GCGAGTAACCTTTCGAGTTTTTGTATCTGAGGAGGAAAGCGTTTACGATTTG 77  
 DB 211 AlaSerAsnValSerIlePheCysValLeuGlnLeuGluSerMetLysLeuProSerLeu 230  
 QY 76 CAGTACGTTTCGCCGCAAGCGCGGTTCGCCAGCTTCGCCCATTTTATGCTGATAAT 17  
 DB 231 ProTyr-----AsnIleAspAlaHisThrLysProThrProAspGly 244  
 QY 16 GATCAC 11  
 DB 245 AspHis 246

## RESULT 6

US-08-399-986B-2  
 ; Sequence 2, Application US/08399986B  
 ; Patent No. 5801041  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Godwin, Andrew K.  
 ; TITLE OF INVENTION: NO. 5801041el Gene Associated with Suppression  
 ; TITLE OF INVENTION: of Tumor Development  
 ; NUMBER OF SEQUENCES: 35  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Darrn, Dorfman, Herrell and Skillman  
 ; STREET: 1601 Market Street  
 ; CITY: Philadelphia  
 ; STATE: PA  
 ; COUNTRY: USA  
 ; ZIP: 19103-2307  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/399,986B  
 ; FILING DATE: 06-MAR-1995  
 ; CLASSIFICATION: 530  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Hagan, Patrick J.  
 ; REGISTRATION NUMBER: 27,643  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (215) 563-4100  
 ; TELEFAX: (215) 563-4044  
 ; INFORMATION FOR SEQ ID NO: 2:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 443 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: not relevant  
 ; TOPOLOGY: not relevant  
 ; MOLECULE TYPE: protein  
 ; HYPOTHEICAL: NO  
 ; ANTI-SENSE: NO  
 US-08-399-986B-2

Alignment Scores:  
 Pred. No.: 20.1 Length: 443  
 Score: 52.50 Matches: 14  
 Percent Similarity: 48.08% Conservative: 11  
 Best Local Similarity: 26.92% Mismatches: 20  
 Query Match: 19.81% Indels: 7  
 DB: 1 Gaps: 1

US-09-856-221-2 (1-154) x US-08-399-986B-2 (1-443)

QY 10 CGTGATCATATCAGCATAAATGCGCAACTGCGTCAACGCGCGCTTCCGCGCGAAA 69  
 Db 265 ArgGluHisTyrAspHisGlnArgMetGlnAlaAlaArgGlnGluAlaIleAlaThrAla 284  
 QY 70 CGTACTGCAATTCGTTAAACCGCTTTGTC-----CTTCCTCAG 108  
 Db 285 ArgSerAlaLysSerTrpGlyLeuIleLeuGlyThrLeuGlyArgGlnGlySerProLys 304  
 QY 109 ATAAACAAAACGCAAGTCTACTGGCAGCGTTA 144  
 Db 305 IleLeuGluHisLeuGluSerArgLeuArgAlaLeu 316

RESULT 7  
 US-08-493-754A-2  
 ; Sequence 2, Application US/08493754A  
 ; Patent No. 5821338  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Godwin, Andrew K.  
 ; TITLE OF INVENTION: No. 5821338el Gene Associated with Suppression  
 ; TITLE OF INVENTION: of Tumor Development  
 ; NUMBER OF SEQUENCES: 35  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Dann, Dorfman, Herrell and Skillman  
 ; STREET: 1601 Market Street  
 ; CITY: Philadelphia  
 ; STATE: PA  
 ; COUNTRY: USA  
 ; ZIP: 19103-2307  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION NUMBER: US/08/493,754A  
 ; FILING DATE: 22-JUN-1995  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Hagan, Patrick J.  
 ; REGISTRATION NUMBER: 27,643  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (215) 563-4100  
 ; TELEFAX: (215) 563-4044  
 ; INFORMATION FOR SEQ ID NO: 2:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 443 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: not relevant  
 ; TOPOLOGY: not relevant  
 ; MOLECULE TYPE: protein  
 ; HYPOTHETICAL: NO  
 ; ANTI-SENSE: NO  
 ; US-08-493-754A-2

Alignment Scores:  
 Pred. No.: 20.1 Length: 443  
 Score: 52.50 Matches: 14  
 Percent Similarity: 48.08% Conservative: 11  
 Best Local Similarity: 26.92% Mismatches: 20  
 Query Match: 19.81% Indels: 7  
 DB: 2 Gaps: 1

US-09-856-221-2 (1-154) x US-08-493-754A-2 (1-443)  
 QY 10 CGTGATCATATCAGCATAAATGCGCAACTGCGTCAACGCGCGCTTCCGCGCGAAA 69  
 Db 265 ArgGluHisTyrAspHisGlnArgMetGlnAlaAlaArgGlnGluAlaIleAlaThrAla 284  
 QY 70 CGTACTGCAATTCGTTAAACCGCTTTGTC-----CTTCCTCAG 108  
 Db 285 ArgSerAlaLysSerTrpGlyLeuIleLeuGlyThrLeuGlyArgGlnGlySerProLys 304  
 QY 109 ATAAACAAAACGCAAGTCTACTGGCAGCGTTA 144

Db 305 IleLeuGluHisLeuGluSerArgLeuArgAlaLeu 316

RESULT 8  
 US-08-887-534A-85  
 ; Sequence 85, Application US/08887534A  
 ; Patent No. 6455323  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Holden, David W.  
 ; TITLE OF INVENTION: ANTI-BACTERIAL METHODS AND MATERIALS  
 ; NUMBER OF SEQUENCES: 106  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
 ; STREET: 233 South Wacker Drive/6300 Sears Tower  
 ; CITY: Chicago  
 ; STATE: Illinois  
 ; COUNTRY: United States of America  
 ; ZIP: 60606  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION NUMBER: US/08/887,534A  
 ; FILING DATE:  
 ; CLASSIFICATION: 536  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Rin-Laures, Li-Hsien  
 ; REGISTRATION NUMBER: 33,547  
 ; REFERENCE/DOCKET NUMBER: 28341/33996  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (312) 474-6300  
 ; TELEFAX: (312) 474-0448  
 ; TELEX: (312) 474-6600  
 ; INFORMATION FOR SEQ ID NO: 85:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1048 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-887-534A-85

Alignment Scores:  
 Pred. No.: 26.2 Length: 1048  
 Score: 52.50 Matches: 14  
 Percent Similarity: 50.00% Conservative: 12  
 Best Local Similarity: 26.92% Mismatches: 23  
 Query Match: 19.81% Indels: 3  
 DB: 4 Gaps: 1

US-09-856-221-2 (1-154) x US-08-887-534A-85 (1-1048)  
 QY 4 ACCATTGCGTATCATATCAGCAT-----AAATGCGCAACTGCGTCAACGCGG 54  
 Db 250 ThrArgGlnAspGluLeuGlnGlnAlaSerArgArgGlnGlnAlaLeuGlnAla 269  
 QY 55 GCCTTCCGCGCAAGTCTACTGCAATTCGTTAAACCGCTTTGTTCTCTCAGATAAAC 114  
 Db 270 LeuAlaGluGluGlyLysAlaGlnProGlnLeuAlaLeuSerLeuAlaGlnProAla 289  
 QY 115 AAAAACTGCAAGTTACTGCGCAGACGTTAGCACAA 150  
 Db 290 ArgAsnLeuArgProHisTrpGluArgIleAlaGlu 301

RESULT 9  
 US-09-248-335-46  
 ; Sequence 46, Application US/09248335  
 ; Patent No. 6096504  
 ; GENERAL INFORMATION:  
 ; APPLICANT: MCGONIGLE, BRIAN  
 ; APPLICANT: O'KEEF, DANIEL  
 ; TITLE OF INVENTION: PLANT GLUTATHIONE-S-TRANSFERASE ENZYMES

```
; FILE REFERENCE: CL-1128-A
; CURRENT APPLICATION NUMBER: US/09/248,335
; CURRENT FILING DATE: 1999-02-10
; EARLIER APPLICATION NUMBER: 08/924,759
; EARLIER FILING DATE: 1997-September-05
; NUMBER OF SEQ ID NOS: 74
; SOFTWARE: Microsoft Word Version 7.0A
; SEQ ID NO 46
; LENGTH: 227
; TYPE: PRT
; ORGANISM: maize
US-09-248-335-46

Alignment Scores:
Pred. No.: 23-2 Length: 227
Score: 51.50 Matches: 11
Percent Similarity: 46.00% Conservative: 12
Best Local Similarity: 22.00% Mismatches: 24
Query Match: 19.29% Indels: 3
DB: 3 Gaps: 1

US-09-856-221-2 (1-154) x US-09-248-335-46 (1-227)
QY 154 GCGTGTGCTGTAACGTCGCGCACTTTCAGTTTGTTCGAGTTTGTTCGAGGAAGAACA 95
Db 171 GlyValAlaLeuLeuAspAlaAspGluPheProAlaLeuCys-----GlnTrpAla 187
QY 94 AAGCGGTTACGAATTCAGTACGTTTCGCGCGCAAGCGCGGTTGACGCGGTTGCC 35
Db 188 ArgAspTyrSerSerGluAlaLeuArgProCysLeuProAspArgAspArgLeuVal 207
QY 34 GCATTTATGCTGATAATGATCAGCAATGG 5
Db 208 AlaTyrPheThrGluAsnLysGluLysTyr 217

RESULT 10
US-09-121-979-4
; Sequence 4, Application US/09121979
; Patent No. 6159709
; GENERAL INFORMATION:
; APPLICANT: Korneluk, Robert G.
; APPLICANT: Holcik, Martin
; APPLICANT: Liston, Peter
; TITLE OF INVENTION: XIAP IRES AND USES THEREOF
; FILE REFERENCE: 07891/021001
; CURRENT APPLICATION NUMBER: US/09/121,979
; CURRENT FILING DATE: 1998-07-24
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 236
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-121-979-4

Alignment Scores:
Pred. No.: 27-9 Length: 236
Score: 51.00 Matches: 10
Percent Similarity: 54.84% Conservative: 7
Best Local Similarity: 32.26% Mismatches: 12
Query Match: 19.25% Indels: 2
DB: 4 Gaps: 1

US-09-856-221-2 (1-154) x US-09-121-979-4 (1-236)
QY 8 TTCGTGATCATTTATCAGCATAAAATCGGCAACTGCGTCAACGCGCGCTTGCGCGCGA 67
Db 196 TyrIleAlaValPheIleProCysGlyHisLeuValThrCysLysGlnCysAlaGlu 215
QY 68 AACGCTACTGCAAAATTCGTTAACGCGTTTGTTC 100
Db 216 AlaVal-----AspArgCysPrometCysSer 224

RESULT 11
US-09-332-319-4
; Sequence 4, Application US/09332319
; Patent No. 6171821
; GENERAL INFORMATION:
; APPLICANT: Korneluk, Robert G.
; APPLICANT: Holcik, Martin
; APPLICANT: Liston, Peter
; TITLE OF INVENTION: XIAP IRES AND USES THEREOF
; FILE REFERENCE: 07891/021002
; CURRENT APPLICATION NUMBER: US/09/332,319
; CURRENT FILING DATE: 1999-06-14
; EARLIER APPLICATION NUMBER: 09/121,979
; EARLIER FILING DATE: 1998-07-24
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 236
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-332-319-4

Alignment Scores:
Pred. No.: 27-9 Length: 236
Score: 51.00 Matches: 10
Percent Similarity: 54.84% Conservative: 7
Best Local Similarity: 32.26% Mismatches: 12
Query Match: 19.25% Indels: 2
DB: 4 Gaps: 1

US-09-856-221-2 (1-154) x US-09-332-319-4 (1-236)
QY 8 TTCGTGATCATTTATCAGCATAAAATCGGCAACTGCGTCAACGCGCGCTTGCGCGCGA 67
Db 196 TyrIleAlaValPheIleProCysGlyHisLeuValThrCysLysGlnCysAlaGlu 215
QY 68 AACGCTACTGCAAAATTCGTTAACGCGTTTGTTC 100
Db 216 AlaVal-----AspArgCysPrometCysSer 224

RESULT 12
US-09-239-867-2
; Sequence 2, Application US/09239867
; Patent No. 6331412
; GENERAL INFORMATION:
; APPLICANT: Robert G. Korneluk et al.
; TITLE OF INVENTION: METHODS AND COMPOUNDS FOR MODULATING
; FILE REFERENCE: 07891/018002
; CURRENT APPLICATION NUMBER: US/09/239,867
; CURRENT FILING DATE: 1999-01-29
; PRIOR APPLICATION NUMBER: 60/073,001
; PRIOR FILING DATE: 1998-01-29
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 236
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-239-867-2

Alignment Scores:
Pred. No.: 27-9 Length: 236
Score: 51.00 Matches: 10
Percent Similarity: 54.84% Conservative: 7
Best Local Similarity: 32.26% Mismatches: 12
Query Match: 19.25% Indels: 2
DB: 4 Gaps: 1

US-09-856-221-2 (1-154) x US-09-239-867-2 (1-236)
QY 8 TTCGTGATCATTTATCAGCATAAAATCGGCAACTGCGTCAACGCGCGCTTGCGCGCGA 67
Db 196 TyrIleAlaValPheIleProCysGlyHisLeuValThrCysLysGlnCysAlaGlu 215
QY 68 AACGCTACTGCAAAATTCGTTAACGCGTTTGTTC 100
Db 216 AlaVal-----AspArgCysPrometCysSer 224
```



```

Best Local Similarity: 30.19%      Mismatches: 28
Query Match: 18.91%      Indels: 4
DB: 2      Gaps: 1

US-09-856-221-2 (1-154) x US-08-673-789-6 (1-984)

QY 152 CGTTGCTAAGCTCTGCCAGTAAGTTTGCAGTTTTT-----TGTTTATCTGAGGA 102
    :||| :|:|:| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 195 LysCysProSerIleValGln-AsnPheAlaValPheProGluThrMetThrGlyAlaG1 214

QY 101 AGGAACAAGCGGTTAACGANTTTCAGTACGTTTCGCCGCGCAGCGCGCGTTGACGC 42
    | :||| :|:| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 214 userThrSerIleuValIleAlaAargGlyThrCysIleProAsnAlaGluGluValAspVa 234

QY 41 AGTTGCCGCATTTTATGCTGATAATGATCACCAGATGG 5
    ||| :|:| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 234 lProIleLysLeuTyrcysAsnGlyAspGlyGluTrp 245
  
```

Search completed: January 15, 2003, 15:57:44  
 Job time : 6.96166 secs





GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: January 15, 2003, 15:46:43 ; Search time 2.88907 Seconds  
(without alignments)  
2119.222 Million cell updates/sec

Title: US-09-856-221-2  
Perfect score: 265  
Sequence: 1 caaacattcgtgatcatta.....gcagacgttagcacaacgcc 154

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 120991 seqs, 19878514 residues

Total number of hits satisfying chosen parameters: 241982

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model  
-O=/cgn2\_1/USPTO\_spool/US09856221/runat\_15012003\_153926\_12555/app.query.fasta\_1.1948  
-DB=Published Applications\_AA -QFMT=Eastan -SUFFIX=rapb -MINMATCH=0.1  
-LOOPCT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US09856221@cgn\_1\_1\_14@runat\_15012003\_153926\_12555  
-NCPU=6 -ICPU=3 -NO\_XLUPY -NO\_WMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG  
-FEU\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications\_AA:

- 1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	126.5	47.7	2504	10	US-09-817-514A-8
2	118.5	44.7	2516	10	US-09-817-514A-2
3	57	21.3	322	9	US-09-966-546-20
4	57	21.3	322	9	US-09-966-545-20
					Sequence 8, Appli
					Sequence 2, Appli
					Sequence 20, Appli

C	5	57	21.3	322	9	US-09-965-212-20	Sequence 20, Appli
C	6	57	21.3	322	9	US-10-174-590-172	Sequence 172, App
C	7	57	21.3	322	9	US-10-176-758-172	Sequence 172, App
C	8	57	21.3	322	12	US-10-052-586-172	Sequence 172, App
C	9	56	21.1	355	10	US-09-741-669-454	Sequence 454, App
C	10	56	21.1	355	10	US-09-815-242-10031	Sequence 10031, A
C	11	54.5	20.4	1232	10	US-09-801-574-46	Sequence 46, Appli
C	12	53.5	20.0	302	10	US-09-903-814-14	Sequence 14, Appli
C	13	53	20.0	215	9	US-09-764-868-1004	Sequence 1004, App
C	14	53	19.9	312	10	US-09-801-368-438	Sequence 438, App
C	15	53	20.0	355	10	US-09-815-242-14078	Sequence 14078, A
C	16	52.5	19.8	1048	10	US-09-741-669-409	Sequence 409, App
C	17	52.5	19.8	1048	10	US-09-815-242-10062	Sequence 10062, A
C	18	52	19.6	566	10	US-09-801-368-418	Sequence 418, App
C	19	52	19.6	780	9	US-09-738-626-4706	Sequence 4706, Ap
C	20	51.5	19.4	344	10	US-09-815-242-13895	Sequence 13895, A
C	21	51.5	19.3	333	10	US-09-815-242-10116	Sequence 10116, A
C	22	51.5	19.4	1046	9	US-10-078-770-186	Sequence 186, App
C	23	51	19.2	334	9	US-09-738-626-5255	Sequence 5255, Ap
C	24	51	19.2	626	9	US-09-738-626-5415	Sequence 5415, Ap
C	25	50.5	19.1	575	10	US-09-970-711-7	Sequence 7, Appli
C	26	50.5	18.9	2001	9	US-10-072-621-8	Sequence 8, Appli
C	27	50.5	18.9	2436	10	US-09-795-693-8	Sequence 63, Appli
C	28	50	18.7	33	10	US-09-908-322-63	Sequence 5290, Ap
C	29	50	18.9	206	10	US-09-815-242-5290	Sequence 12240, A
C	30	50	18.9	207	10	US-09-815-242-12240	Sequence 12758, A
C	31	50	18.9	207	10	US-09-815-242-12758	Sequence 123, App
C	32	50	18.9	419	10	US-09-894-018-123	Sequence 18, Appli
C	33	50	18.7	647	10	US-09-725-735A-18	Sequence 41723, A
C	34	49.5	18.5	62	10	US-09-863-693-23	Sequence 46739, A
C	35	49.5	18.7	148	10	US-09-864-761-41723	Sequence 12, Appli
C	36	49.5	18.7	148	10	US-09-864-761-46739	Sequence 12, Appli
C	37	49.5	18.7	374	10	US-09-853-625B-12	Sequence 8, Appli
C	38	49.5	18.5	598	10	US-09-875-811-12	Sequence 4, Appli
C	39	49.5	18.5	605	10	US-09-875-811-8	Sequence 10, Appli
C	40	49.5	18.5	621	10	US-09-875-811-4	Sequence 2, Appli
C	41	49.5	18.5	656	10	US-09-942-446-2	Sequence 6, Appli
C	42	49.5	18.5	656	10	US-09-875-811-6	Sequence 2, Appli
C	43	49.5	18.5	663	10	US-09-875-811-2	Sequence 10, Appli
C	44	49.5	18.5	679	10	US-09-858-525A-10	Sequence 10, Appli
C	45	49.5	18.7	836	9		

ALIGNMENTS

RESULT 1  
US-09-817-514A-8  
; Sequence 8, Application US/09817514A  
; Patent No. US20020078478A1  
; GENERAL INFORMATION:  
; APPLICANT: Iffrench-Constant, Richard  
; APPLICANT: Bowen, David  
; APPLICANT: Rocheleau, Thomas  
; APPLICANT: Waterfield, Nicholas  
; TITLE OF INVENTION: DNA SEQUENCES FROM PHOTORHABDUS LUMINESCENS  
; FILE REFERENCE: 61645  
; CURRENT APPLICATION NUMBER: US/09/817,514A  
; PRIOR FILING DATE: 2000-03-26  
; PRIOR APPLICATION NUMBER: US 60/191806  
; PRIOR FILING DATE: 2000-03-24  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 8  
; TYPE: PRT  
; ORGANISM: Photorhabdus luminescens  
US-09-817-514A-8

Alignment Scores:  
Pred. No.: 2.47e-10  
Score: 126.50  
Percent Similarity: 64.71%  
Best Local Similarity: 54.90%  
Length: 2504  
Matches: 28  
Conservative: 5  
Mismatches: 17

```
Query Match: 47.74% Indels: 1
DB: 10 Gaps: 1
US-09-856-221-2 (1-154) x US-09-817-514A-8 (1-2504)
QY 1 CAACACATTCGTGATCATTTATCAGCATAAATGCGCACTGCGTCAACGCGCGGCTTG 60
Db 1911 GlnGlnValArgGlnGlnValLeuThrGlnLeuArg--LeuAsnSerArgValLysThr 1929
QY 61 CCGCGCAAGAGTACTCAAAATTCGTTAAACCGCTTTGTTCCTTCCTCAGATAACAAAAA 120
Db 1930 ProLeuLeuGlyThrAlaAsnSerLeuThrAlaLeuPheLeuProGlnGluAsnSerLys 1949
QY 121 CTGCAAGAGTACGCGAGACGTTAGCACACGC 153
Db 1950 LeuLysGlyTyrTrpArgThrLeuAlaGlnArg 1960
RESULT 2
US-09-817-514A-2
; Sequence 2, Application US/09817514A
; Patent No. US20020078478A1
; GENERAL INFORMATION:
; APPLICANT: ffrench-Constant, Richard
; APPLICANT: Bowen, David
; APPLICANT: Rocheleau, Thomas
; APPLICANT: Waterfield, Nicholas
; TITLE OF INVENTION: DNA SEQUENCES FROM PHOTORHABDUS LUMINESCENS
; FILE REFERENCE: 61645
; CURRENT APPLICATION NUMBER: US/09/817,514A
; CURRENT FILING DATE: 2000-03-26
; PRIOR APPLICATION NUMBER: US 60/191806
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 2516
; TYPE: PRT
; ORGANISM: Photorhabdus luminescens
US-09-817-514A-2
Alignment Scores:
Pred. No.: 3,77e-09 Length: 2516
Score: 118.50 Matches: 26
Percent Similarity: 66.04% Conservative: 9
Best Local Similarity: 49.06% Mismatches: 15
Query Match: 44.72% Indels: 3
DB: 10 Gaps: 1
US-09-856-221-2 (1-154) x US-09-817-514A-2 (1-2516)
QY 4 ACCATTCGTGATCATTTATCAGCATAAATGCGCAACTGCGTCAACGCGCGCTTGCCG 63
Db 1914 ThrThrGlnAsnAlaHisAspSerAlaIleValAlaLeuArgGlnAsnIleProThrPro 1933
QY 64 CGGAAA-----CGTACTGCAAAATTCGTTAAACCGCTTTGTTCCTTCCTCAGATAAAC 114
Db 1934 AlaProLeuSerLeuArgSerAlaAsnThrLeuThrAspLeuPheLeuProGlnIleAsn 1953
QY 115 AAAAATCGCAAGAGTACTGCGAGACGTTAGCACACGC 153
Db 1954 GluValMetMetAsnTyrTrpGlnThrLeuAlaGlnArg 1966
RESULT 3
US-09-966-546-20
; Sequence 20, Application US/09966546
; Patent No. US20020168716A1
; GENERAL INFORMATION:
; APPLICANT: Fernandes, Elma
; APPLICANT: Vernet, Corine
; APPLICANT: Shmukets, Richard A.
; TITLE OF INVENTION: No. US20020168716A1 Human Proteins and Polynucleotides Encoding
; FILE REFERENCE: Cura-46 (15966-546)
```

```
; CURRENT APPLICATION NUMBER: US/09/966,546
; CURRENT FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 09/544,511
; PRIOR FILING DATE: 2000-04-06
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
; LENGTH: 322
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-966-546-20
Alignment Scores:
Pred. No.: 2,76 Length: 322
Score: 57.00 Matches: 15
Percent Similarity: 55.00% Conservative: 7
Best Local Similarity: 37.50% Mismatches: 18
Query Match: 21.35% Indels: 0
DB: 9 Gaps: 0
US-09-856-221-2 (1-154) x US-09-966-546-20 (1-322)
QY 145 CTAACGCTGCCAGTAACCTTTGCGAGTTTGTATTCTGAGGAGGAACAAGCGGTTA 86
Db 103 IleThrPheAlaCysTyrAlaAlaLeuPheCysLeuSerAlaSerIleIleTyrProThr 122
QY 85 ACGAATTTGCAGTACGTTTTCGCCGCGCAAGCGCGTTTGACGCAGTTGCCGCAATTTAT 26
Db 123 ThrTyrValGlnPheLeuSerHisGlyArgSerArgAspHisAlaIleAlaThrPhe 142
RESULT 4
US-09-966-545-20
; Sequence 20, Application US/09966545
; Patent No. US20020172999A1
; GENERAL INFORMATION:
; APPLICANT: Fernandes, Elma
; APPLICANT: Vernet, Corine
; APPLICANT: Shmukets, Richard A.
; TITLE OF INVENTION: No. US20020172999A1 Human Proteins and Polynucleotides Encod
; FILE REFERENCE: Cura-46 (15966-546)
; CURRENT APPLICATION NUMBER: US/09/966,545
; CURRENT FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 09/544,511
; PRIOR FILING DATE: 2000-04-06
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
; LENGTH: 322
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-966-545-20
Alignment Scores:
Pred. No.: 2,76 Length: 322
Score: 57.00 Matches: 15
Percent Similarity: 55.00% Conservative: 7
Best Local Similarity: 37.50% Mismatches: 18
Query Match: 21.35% Indels: 0
DB: 9 Gaps: 0
US-09-856-221-2 (1-154) x US-09-966-545-20 (1-322)
QY 145 CTAACGCTGCCAGTAACCTTTGCGAGTTTGTATTCTGAGGAGGAACAAGCGGTTA 86
Db 103 IleThrPheAlaCysTyrAlaAlaLeuPheCysLeuSerAlaSerIleIleTyrProThr 122
QY 85 ACGAATTTGCAGTACGTTTTCGCCGCGCAAGCGCGCTTGACGCAGTTGCCGCAATTTAT 26
Db 123 ThrTyrValGlnPheLeuSerHisGlyArgSerArgAspHisAlaIleAlaThrPhe 142
RESULT 5
US-09-965-212-20
```

```

; Sequence 20, Application US/09965212
; Publication No. US2003003462A1
; GENERAL INFORMATION:
; APPLICANT: Fernandes, Elma
; APPLICANT: Vernet, Corine
; APPLICANT: Shimkets, Richard A.
; TITLE OF INVENTION: Them
; FILE REFERENCE: Cura-46 (15966-546)
; CURRENT APPLICATION NUMBER: US/09/965,212
; CURRENT FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US/09/544,511
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: USSN 60/128,514
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
; LENGTH: 322
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-965-212-20

Alignment Scores:
Pred. No.: 2,76 Length: 322
Score: 57.00 Matches: 15
Percent Similarity: 55.00% Conservative: 7
Best Local Similarity: 37.50% Mismatches: 18
Query Match: 21.35% Indels: 0
DB: Gaps: 0

US-09-856-221-2 (1-154) x US-09-965-212-20 (1-322)
QY 145 CTAACGCTGCCAGTAACTTTGCGAGTTTGTATCTGAGGAGGACAAACCGGTTA 86
Db 103 IleThrPheAlaCysTyraAlaLeuPheCysLeuSerAlaSerIleIleTyProThr 122
QY 85 ACAGATTTGCGAGTACGTTTCGCCGCGGCAAGCCGCGTTGACGCGAGTTGCCGATTTTAT 26
Db 123 ThrTyValGlnPheLeuSerHisGlyArgSerArgAspHisAlaIleAlaThrPhe 142

RESULT 6
US-10-174-590-172
; Sequence 172, Application US/10174590
; Publication No. US20030008352A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C42
; CURRENT APPLICATION NUMBER: US/10/174,590
; CURRENT FILING DATE: 2002-06-18
; Prior application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 172
; LENGTH: 322
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-174-590-172

Alignment Scores:
Pred. No.: 2,76 Length: 322
Score: 57.00 Matches: 15
Percent Similarity: 55.00% Conservative: 7
Best Local Similarity: 37.50% Mismatches: 18
Query Match: 21.35% Indels: 0
DB: Gaps: 0

US-09-856-221-2 (1-154) x US-10-174-590-172 (1-322)
QY 145 CTAACGCTGCCAGTAACTTTGCGAGTTTGTATCTGAGGAGGACAAACCGGTTA 86
Db 103 IleThrPheAlaCysTyraAlaLeuPheCysLeuSerAlaSerIleIleTyProThr 122
QY 85 ACAGATTTGCGAGTACGTTTCGCCGCGGCAAGCCGCGTTGACGCGAGTTGCCGATTTTAT 26
Db 123 ThrTyValGlnPheLeuSerHisGlyArgSerArgAspHisAlaIleAlaThrPhe 142

RESULT 7
US-10-176-758-172
; Sequence 172, Application US/10176758
; Publication No. US20030008353A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C104
; CURRENT APPLICATION NUMBER: US/10/176,758
; CURRENT FILING DATE: 2002-06-21
; Prior application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 172
; LENGTH: 322
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-176-758-172

Alignment Scores:
Pred. No.: 2,76 Length: 322
Score: 57.00 Matches: 15
Percent Similarity: 55.00% Conservative: 7
Best Local Similarity: 37.50% Mismatches: 18
Query Match: 21.35% Indels: 0
DB: Gaps: 0

US-09-856-221-2 (1-154) x US-10-176-758-172 (1-322)
QY 145 CTAACGCTGCCAGTAACTTTGCGAGTTTGTATCTGAGGAGGACAAACCGGTTA 86
Db 103 IleThrPheAlaCysTyraAlaLeuPheCysLeuSerAlaSerIleIleTyProThr 122
QY 85 ACAGATTTGCGAGTACGTTTCGCCGCGGCAAGCCGCGTTGACGCGAGTTGCCGATTTTAT 26
Db 123 ThrTyValGlnPheLeuSerHisGlyArgSerArgAspHisAlaIleAlaThrPhe 142

RESULT 8
US-10-052-586-172
; Sequence 172, Application US/10052586
; Patent No. US20020127584A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.

```

APPLICANT: Pan,James  
APPLICANT: Smith,Victoria  
APPLICANT: Watanabe,Colin K.  
APPLICANT: Wood,William I.  
APPLICANT: Zhang,Zemin  
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
FILE REFERENCE: P343081C1  
CURRENT APPLICATION NUMBER: US/10/052,586  
CURRENT FILING DATE: 2002-01-15  
PRIOR APPLICATION NUMBER: 60/059263  
PRIOR FILING DATE: 1997-09-18  
PRIOR APPLICATION NUMBER: 60/059266  
PRIOR FILING DATE: 1997-09-18  
PRIOR APPLICATION NUMBER: 60/062250  
PRIOR FILING DATE: 1997-10-17  
PRIOR APPLICATION NUMBER: 60/063120  
PRIOR FILING DATE: 1997-10-24  
PRIOR APPLICATION NUMBER: 60/063121  
PRIOR FILING DATE: 1997-10-24  
PRIOR APPLICATION NUMBER: 60/063486  
PRIOR FILING DATE: 1997-10-21  
PRIOR APPLICATION NUMBER: 60/063540  
PRIOR FILING DATE: 1997-10-28  
PRIOR APPLICATION NUMBER: 60/063541  
PRIOR FILING DATE: 1997-10-28  
PRIOR APPLICATION NUMBER: 60/063544  
PRIOR FILING DATE: 1997-10-28  
PRIOR APPLICATION NUMBER: 60/063564  
PRIOR FILING DATE: 1997-10-28  
PRIOR APPLICATION NUMBER: 60/063734  
PRIOR FILING DATE: 1997-10-29  
PRIOR APPLICATION NUMBER: 60/063870  
PRIOR FILING DATE: 1997-10-31  
PRIOR APPLICATION NUMBER: 60/064103  
PRIOR FILING DATE: 1997-10-31  
PRIOR APPLICATION NUMBER: 60/065311  
PRIOR FILING DATE: 1997-11-13  
PRIOR APPLICATION NUMBER: 60/066120  
PRIOR FILING DATE: 1997-11-21  
PRIOR APPLICATION NUMBER: 60/066466  
PRIOR FILING DATE: 1997-11-24  
PRIOR APPLICATION NUMBER: 60/066772  
PRIOR FILING DATE: 1997-11-24  
PRIOR APPLICATION NUMBER: 60/069335  
PRIOR FILING DATE: 1997-12-11  
PRIOR APPLICATION NUMBER: 60/069425  
PRIOR FILING DATE: 1997-12-12  
PRIOR APPLICATION NUMBER: 60/069870  
PRIOR FILING DATE: 1997-12-17  
PRIOR APPLICATION NUMBER: 60/068017  
PRIOR FILING DATE: 1997-12-18  
PRIOR APPLICATION NUMBER: 60/077450  
PRIOR FILING DATE: 1998-03-10  
PRIOR APPLICATION NUMBER: 60/077632  
PRIOR FILING DATE: 1998-03-11  
PRIOR APPLICATION NUMBER: 60/077649  
PRIOR FILING DATE: 1998-03-11  
PRIOR APPLICATION NUMBER: 60/078886  
PRIOR FILING DATE: 1998-03-20  
PRIOR APPLICATION NUMBER: 60/078939  
PRIOR FILING DATE: 1998-03-20  
PRIOR APPLICATION NUMBER: 60/079664  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/079786  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/080107  
PRIOR FILING DATE: 1998-03-31  
PRIOR APPLICATION NUMBER: 60/080194  
PRIOR FILING DATE: 1998-03-31  
PRIOR APPLICATION NUMBER: 60/080327  
PRIOR FILING DATE: 1998-04-01  
PRIOR APPLICATION NUMBER: 60/080333  
PRIOR FILING DATE: 1998-04-01  
PRIOR APPLICATION NUMBER: 60/081049  
PRIOR FILING DATE: 1998-04-08  
PRIOR APPLICATION NUMBER: 60/081070  
PRIOR FILING DATE: 1998-04-08  
PRIOR APPLICATION NUMBER: 60/081195  
PRIOR FILING DATE: 1998-04-09  
PRIOR APPLICATION NUMBER: 60/081838  
PRIOR FILING DATE: 1998-04-15  
PRIOR APPLICATION NUMBER: 60/082568  
PRIOR FILING DATE: 1998-04-21  
PRIOR APPLICATION NUMBER: 60/082569  
PRIOR FILING DATE: 1998-04-21  
PRIOR APPLICATION NUMBER: 60/082704  
PRIOR FILING DATE: 1998-04-22  
PRIOR APPLICATION NUMBER: 60/082797  
PRIOR FILING DATE: 1998-04-22  
PRIOR APPLICATION NUMBER: 60/083322  
PRIOR FILING DATE: 1998-04-28  
PRIOR APPLICATION NUMBER: 60/083495  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083496  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083499  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083559  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/084366  
PRIOR FILING DATE: 1998-05-05  
PRIOR APPLICATION NUMBER: 60/084414  
PRIOR FILING DATE: 1998-05-06  
PRIOR APPLICATION NUMBER: 60/084639  
PRIOR FILING DATE: 1998-05-07  
PRIOR APPLICATION NUMBER: 60/084640  
PRIOR FILING DATE: 1998-05-07  
PRIOR APPLICATION NUMBER: 60/084643  
PRIOR FILING DATE: 1998-05-07  
PRIOR APPLICATION NUMBER: 60/085573  
PRIOR FILING DATE: 1998-05-15  
PRIOR APPLICATION NUMBER: 60/085579  
PRIOR FILING DATE: 1998-05-15  
PRIOR APPLICATION NUMBER: 60/085580  
PRIOR FILING DATE: 1998-05-15  
PRIOR APPLICATION NUMBER: 60/085582  
PRIOR FILING DATE: 1998-05-15  
PRIOR APPLICATION NUMBER: 60/085700  
PRIOR FILING DATE: 1998-05-15  
PRIOR APPLICATION NUMBER: 60/086023  
PRIOR FILING DATE: 1998-05-18  
PRIOR APPLICATION NUMBER: 60/086392  
PRIOR FILING DATE: 1998-05-22  
PRIOR APPLICATION NUMBER: 60/086486  
PRIOR FILING DATE: 1998-05-22  
PRIOR APPLICATION NUMBER: 60/087098  
PRIOR FILING DATE: 1998-05-28  
PRIOR APPLICATION NUMBER: 60/087208  
PRIOR FILING DATE: 1998-05-28  
PRIOR APPLICATION NUMBER: 60/087609  
PRIOR FILING DATE: 1998-06-02  
PRIOR APPLICATION NUMBER: 60/087759  
PRIOR FILING DATE: 1998-06-02  
PRIOR APPLICATION NUMBER: 60/087827  
PRIOR FILING DATE: 1998-06-03  
PRIOR APPLICATION NUMBER: 60/088025  
PRIOR FILING DATE: 1998-06-04  
PRIOR APPLICATION NUMBER: 60/088028  
PRIOR FILING DATE: 1998-06-04  
PRIOR APPLICATION NUMBER: 60/088029  
PRIOR FILING DATE: 1998-06-04  
PRIOR APPLICATION NUMBER: 60/088033  
PRIOR FILING DATE: 1998-06-04  
PRIOR APPLICATION NUMBER: 60/088167  
PRIOR FILING DATE: 1998-06-05

; TITLE OF INVENTION: Genes identified as required for

```

; TITLE OF INVENTION: proliferation of E. coli
; FILE REFERENCE: ELITRA.009A
; CURRENT APPLICATION NUMBER: US/09/741,669
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: US 60/173005
; PRIOR FILING DATE: 1999-12-23
; NUMBER OF SEQ ID NOS: 481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 454
; LENGTH: 355
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-669-454

Alignment Scores:
Pred. No.: 3.98 Length: 355
Score: 56.00 Matches: 18
Percent Similarity: 37.14% Conservative: 8
Best Local Similarity: 25.71% Mismatches: 22
Query Match: 21.13% Indels: 22
DB: 10 Gaps: 2

US-09-856-22I-2 (1-154) x US-09-741-669-454 (1-355)

QY 1 CAACACCATCGTGATCATTTAT-----CAGCATAAATGCGCAACTG 42
|||||:|||||:|||||:|||||:|||||:
Db 224 GlnSerValGluGlnAlaTyRAlaGluAlaGlyGlnProGlnHisLysValThrGluPhe 263
|||||:|||||:|||||:|||||:|||||:
QY 43 -----CGTCAACGCGCG 54
|||||:|||||:|||||:|||||:
Db 244 IleAspMetAlaAlaAlaTyRAlaTrpAlaAspValValValCysArgSerGlyAla 263
QY 55 GCCTTGCGCGGCAACGCTACTGCAATTCGTTAACCGCTTTGCTTCCTTCCTCAGATAAAC 114
|||||:|||||:|||||:|||||:|||||:
Db 264 LeuThrValSerGluIleAlaAlaAlaGlyLeuProAlaLeuPheValProPheGlnHis 283
|||||:|||||:|||||:|||||:|||||:
QY 115 AAAAACTGCAAACTTACTGCGACAGCTTA 144
|||||:|||||:|||||:|||||:
Db 284 LysAspArgGlnGlnTyrTrpAsnAlaLeu 293

RESULT 10
US-09-815-242-10031
; Sequence 10031, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; PROKARYOTES
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110

```

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10031
; LENGTH: 355
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-815-242-10031
```

```
Alignment Scores:
Pred. No.: 3.98 Length: 355
Score: 56.00 Matches: 18
Percent Similarity: 37.14% Conservative: 8
Best Local Similarity: 25.71% Mismatches: 22
Query Match: 21.13% Indels: 22
DB: 10 Gaps: 2
```

US-09-856-221-2 (1-154) x US-09-815-242-10031 (1-355)

```
QY 1 CAACACATTCGTGATCATTTAT-----CAGCATAAAATCGCGCAACTG 42
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 224 GlnSerValGluGlnAlaTyAlaGluAlaGlyGlnProGlnHisLysValThrGluPhe 243
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 43 -----CGTCACGCGCG 54
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 244 IleAspAspMetAlaAlaAlaTyAlaTrpAlaAspValValValCysArgSerGlyAla 263
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 55 GCCTTGGCGGCAACGTAAGTCTGCAAAATTCGTTAAACGCTTTGTTCTTCCTCAGATAAAC 114
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 264 LeuThrValSerGluIleAlaAlaAlaGlyLeuProAlaLeuPheValProPheGlnHis 283
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 115 AAAAAAGCTGCAAGTTACTGTCACAGCTTGA 144
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 284 LysAspArgGlnGlnTyTrpAsnAlaLeu 293
```

RESULT 11

```
US-09-801-574-46
; Sequence 46, Application US/09801574
; Patent No. US20020081592A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Peijiang Jeremy
; APPLICANT: Page, David C.
; TITLE OF INVENTION: reproduction-specific genes
; FILE REFERENCE: 0399.2007-002
; CURRENT APPLICATION NUMBER: US/09/801,574
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: 60/187,518
; PRIOR FILING DATE: 2000-03-07
; PRIOR APPLICATION NUMBER: 60/261,557
; PRIOR FILING DATE: 2001-01-12
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 46
; LENGTH: 1232
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-801-574-46
```

```
Alignment Scores:
Pred. No.: 9.2 Length: 1232
Score: 54.50 Matches: 13
Percent Similarity: 66.67% Conservative: 9
Best Local Similarity: 39.39% Mismatches: 10
Query Match: 20.41% Indels: 1
DB: 10 Gaps: 1
```

US-09-856-221-2 (1-154) x US-09-801-574-46 (1-1232)

```
QY 154 GCGTGTGCTAACCTCTGCCAGTAAGTTGTCAGTTTGTATCTCAGCAAGGAACA 95
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 147 GlyPheIleHisArgSerLeuSerTyAlaValHisIleValSer---AlaGlyGlu 165
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 94 AAGCGGTTAAGCAATTTGACGTACGTTTCGCGCGCAAGG 56
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 166 AlaArgLeuThrAsnLeuGluTyrLeuThrGluSerGln 178
```

RESULT 12

```
US-09-903-814-14
; Sequence 14, Application US/09903814
; Patent No. US20020102689A1
; GENERAL INFORMATION:
; APPLICANT: Falco, Carl
; APPLICANT: Famodu, Layo O.
; APPLICANT: Orozco, Buddy
; APPLICANT: Rafalski, Antoni
; APPLICANT: Thorpe, Cathy
; TITLE OF INVENTION: Tetrahydrofolate Metabolism Enzymes
; FILE REFERENCE: BB-1179-A
; CURRENT APPLICATION NUMBER: US/09/903,814
; CURRENT FILING DATE: 2001-07-12
; PRIOR APPLICATION NUMBER: 60/092,869
; PRIOR FILING DATE: 1998-07-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 14
; LENGTH: 302
; TYPE: PRT
; ORGANISM: Triticum aestivum
US-09-903-814-14
```

```
Alignment Scores:
Pred. No.: 8.93 Length: 302
Score: 53.50 Matches: 14
Percent Similarity: 47.73% Conservative: 7
Best Local Similarity: 31.82% Mismatches: 22
Query Match: 20.04% Indels: 1
DB: 10 Gaps: 1
```

US-09-856-221-2 (1-154) x US-09-903-814-14 (1-302)

```
QY 145 CTAACGCTCGCCAGTAACCTTTCGAGTTTGTGTTATCTCAGCAAGGAACAAAGCGTTA 86
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 200 LeuValGluSerTyAsnValProIlePheCysLysAspValProThrValArgArgIle 219
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 85 ACG---AATTTGACGTACGTTTCGCGCGCAAGCGCGCTTCGACGAGTTCGCGCATTT 29
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 220 ThrArgArgValThrGlySerGlyLeuProSerValGlnAlaLeuAlaLeuPhe 239
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 28 TATGCTGATTAAT 17
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 240 HisGlyAspAsn 243
```

RESULT 13

```
US-09-764-868-1004
; Sequence 1004, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1004
; LENGTH: 215
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-868-1004
```

```
Alignment Scores:
Pred. No.: 9.68 Length: 215
Score: 53.00 Matches: 12
Percent Similarity: 58.62% Conservative: 5
Best Local Similarity: 41.38% Mismatches: 12
Query Match: 20.00% Indels: 0
DB: 9 Gaps: 0
```

```

US-09-856-221-2 (1-154) x US-09-764-868-1004 (1-215)

QY 58 TTGCGCGGGAACGCTACTGACAAATGCTTAACGCGTTTGTCTCTCTCCTCAGATAACAAA 117
||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
Db 183 LeuProGlyLysLysThrPheAsnAsnMetAspArgPheLeuGluLysArgLys 202
||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||

QY 118 AAACCTGCAAAAGTTACTGCGCAGACGTTA 144
||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
Db 203 AspLeuAsnAlaTyrLeuGlnLeuLeu 211
||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||

RESULT 14
US-09-801-368-438
; Sequence 438, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Cali, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: No. US20020128250A1man, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801,368
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 438
; LENGTH: 312
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-801-368-438

Alignment Scores:
Pred. No.: 10-7 Length: 312
Score: 53.00 Matches: 12
Percent Similarity: 56.67% Conservative: 5
Best Local Similarity: 40.00% Mismatches: 13
Query Match: 19.85% Indels: 0
DB: 10 Gaps: 0

US-09-856-221-2 (1-154) x US-09-801-368-438 (1-312)
QY 133 AGTAACCTTTGACGTTTTTTTGTATCTGAGCAAGGACAAACGGTTAACGAATTTGCAG 74
||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
Db 273 SerAsnPheLysIlePheThrLeuProGluAspPheLysThrIleSerAsnLeuSer 292
||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||

QY 73 TAGCTTTTCCGCGGCAAGCGCGCGCTTGAC 44
||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
Db 293 LysValHisGlyThrLysArgValValAsp 302
||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||

RESULT 15
US-09-815-242-14078
; Sequence 14078, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.

```











US-09-856-221-2 (1-154) x T32189 (1-455)

QY 29 AAATGGCGCAACTGGTCAACGCGCGCTTCGCGGGAACGACTACTGCAAAATCGTTAA 88

Db 85 AtgCysGlySerAlaAsnPhelLysProCysArgAsnMet----- 99

QY 89 CCGCTTTGTTCTTCCTCAGAA---TAAACAAAATACTGCAAA 127

Db 100 ---AsnCysGluPhelLysAsnGlyTyrPheAsnCysLys 112

RESULT 9

F83781 transposase (08) / ABC transporter (ATP-binding protein) BH1054 [imported] - Bacillus ha

C:Species: Bacillus halodurans

C:Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001

C:Accession: F83781

R:Ikami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hirai

Nucleic Acids Res. 28, 4317-4331, 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and

A:Reference number: A83650; MUID:20512582; PMID:11058132

A:Accession: F83781

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-901 <STO>

A:Cross-references: GB:AP001510; GB:BA000004; NID:gl0173440; PIDN:BA04773.1; GSPDB:GN00

A:Experimental source: strain C-125

C:Genetics:

C:Gene: BH1054

Alignment Scores:

Pred. No.: 14 Length: 901

Score: 59.50 Matches: 13

Percent Similarity: 60.78% Conservative: 18

Best Local Similarity: 25.49% Mismatches: 19

Query Match: 22.45% Indels: 1

DB: 2 Gaps: 1

US-09-856-221-2 (1-154) x F83781 (1-901)

QY 1 CAAACCATTCGTGATCATTATCAGCAATAAATCGGCAACTGCGTCAACGCGCGCTTG 60

Db 223 GluGlnIleLysGluThrAlaAspHisLysGluLysGlnAlaGlnLysAlaLeuLys 242

QY 61 CCGCGGAACGACTGCGAAATTCGTTAAACGCGCTTTCCTCCTCAGATAAACAATAA 120

Db 243 ProLeuArgLysIleTyrArgThrLeuThrThrAspThrLeuProArgLeu---ArgLys 261

QY 121 CTGCAAGTTACTGGCAGACGCTAGCACACGC 153

Db 262 TyrGluThrSerPheArgLeuGlyGluArg 272

RESULT 10

AD2325 hypothetical protein all4155 [imported] - Nostoc sp. (strain PCC 7120)

C:Species: Nostoc sp.

A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 30-Jun-2002

C:Accession: AD2325

R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriuguhi

Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S

DNA Res. 8, 205-213, 2001

A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana

A:Reference number: AB1807; MUID:21595285; PMID:11759840

A:Accession: AD2325

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-278 <KUR>

A:Cross-references: GB:BA000019; PIDN:BA075854.1; PID:g17133290; GSPDB:GN00179

A:Experimental source: strain PCC 7120

C:Genetics:

A:Gene: all4155

Alignment Scores:

Pred. No.: 15.4 Length: 278

Score: 59.00 Matches: 21

Percent Similarity: 52.17% Conservative: 3

Best Local Similarity: 45.65% Mismatches: 14

Query Match: 22.26% Indels: 8

DB: 2 Gaps: 3

US-09-856-221-2 (1-154) x AD2325 (1-278)

QY 7 ATTCGTGATCATATACGCAATAAATGCGGCAACTGCTCAACGCGCGCTTGCCTGCGGCG 66

Db 170 IleArgAsp-----LeuArgGlnLeuArg---AsnAlaLysIleGluThr 183

QY 67 AAACGTACTGCAAAATTCGTTAAACGCGCTTGTTCCTTCCT-----CAGATAAACAATAA 120

Db 184 LysIleThrAlaLysSerGluThrGlnGlnIleProArgThrGlnSerAsnIleLys 203

QY 121 CTGCAAGTTACTGCGCAG 138

Db 204 LeuAspSerCysTrpGln 209

RESULT 11

B71361

probable thiamin ABC transporter, thiamin-binding periplasmic protein - syphilis spir

C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)

C:Date: 24-Jul-1998 #sequence\_revision 24-Jul-1998 #text\_change 03-Jun-2002

C:Accession: B71361

R:Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; G

rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; M

they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.

Science 281, 375-388, 1998

A:Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.

A:Reference number: A71250; MUID:98332770; PMID:9665876

A:Accession: B71361

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-335 <COL>

A:Cross-references: GB:AE001199; GB:AE000520; NID:g3322402; PIDN:AAC65133.1; PID:g332

A:Experimental source: strain Nichols

C:Genetics:

C:Gene: Tp0144

C:Superfamily: sfua protein

Alignment Scores:

Pred. No.: 15.5 Length: 335

Score: 59.00 Matches: 16

Percent Similarity: 52.63% Conservative: 4

Best Local Similarity: 42.11% Mismatches: 18

Query Match: 22.10% Indels: 0

DB: 2 Gaps: 0

US-09-856-221-2 (1-154) x B71361 (1-335)

QY 136 GCCAGTACTTGCGATTTTGTATCTGAGGAGGAGCAACAGCGTTACGATTTTG 77

Db 271 AlaThrGlnPheIleAspPheLeuSerThrGluAlaGlnGluLeuLeuProLeuThr 290

QY 76 CAGTACGTTTCGCGGCAAGCGCGCTTGACGCGATTCGCGCATTTTATGCT 23

Db 291 GlnTrpMetTyrProAlaAsnProAspValProLeuProAlaAlaPheSerAla 308

RESULT 12

C87664

methyl-accepting chemotaxis protein McpH [imported] - Caulobacter crescentus

C:Species: Caulobacter crescentus

C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 01-Mar-2002

C:Accession: C87664

R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg,

B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Ko

n, J.; Emolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C

Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A:Title: Complete Genome Sequence of Caulobacter crescentus.



```

DB:          2          Gaps:          2
US-09-856-221-2 (1-154) x G71006 (1-347)
QY 136 GCCAGTAACCTTTGCGAGTTTCTGTTATCTGAGGAAGGAACAAGCGGTTAAGCAATTG 77
    |||          |||          |||          |||          |||          |||
Db 268 AlaGluArgPheValGluPheLeuIleSerGlnLysAlaGlnGluLysLeuProThrThr 287
    |||          |||          |||          |||          |||          |||
QY 76  CAGTACGTTTCGCCGCA-----AGCCCG 53
    |||          |||          |||          |||          |||          |||
Db 288 GlnTrpMetPheProAlaAsnLysLeuProGluValTyrLysTyrAsnPro 307
    |||          |||          |||          |||          |||          |||
QY 52  CGCGTTGACGAGTTGCCGCTTTTATGCTGAT-----AATGATCAC 11
    |||          |||          |||          |||          |||          |||
Db 308 LysValAspPheSerLysAlaValTyrLeuAspProLysLeuIleGlnGluAsnTyrGlu 327
    |||          |||          |||          |||          |||          |||
QY 10  GAATGG 5
    |||          |||          |||          |||          |||          |||
Db 328 LysTrp 329
    |||          |||          |||          |||          |||          |||

```

Search completed: January 15, 2003, 15:56:19  
Job time : 11.0971 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: January 15, 2003, 15:39:33 ; Search time 3.64274 Seconds  
(without alignments)  
3506.895 Million cell updates/sec

Title: US-09-856-221-2  
Perfect score: 265  
Sequence: 1 caaacattcgtgacattt.....gcagagcttagcacacgccc 154

Scoring table: BL0SUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 225784

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-O/cgpn2\_1/USPTO.spool/US09856221/runat\_15012003\_153924\_12445/app\_query.fasta\_1.1948  
-DB=SwissProt40 -QFMT=fastan -SUFFIX=rsp -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=DILTS -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40 -cdi -LIST=45  
-DOLALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USRF=US09856221@cgn\_1\_1\_38@runat\_15012003\_153924\_12445 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAPP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	68	25.7	1718	1 RRPO_SHVX	Q04575 shallot vir
2	63.5	23.8	1720	1 FTSH_CHLVU	P56369 chlorella v
3	60.5	22.8	477	1 Y264_SYNY3	P73436 synchocyst
4	59.5	22.5	455	1 NH55_CAEEL	O16962 caenorhabdi
5	58.5	21.9	490	1 MET3_SCHPO	P78937 schizosacch
6	58	21.9	694	1 LCF3_YEAST	P39002 saccharomyc
7	58	21.9	767	1 HYFE_SYNY3	O55638 synchocyst
8	57.5	21.5	87	1 Y07A_BPT4	P39226 bacterioph
9	57	21.3	322	1 MYDM_HUMAN	O96897 homo sapien
10	56.5	21.2	984	1 EPBL_CHICK	Q07494 gallus gall
11	56	21.1	354	1 MURG_ECOLI	O6x98 escherichia
12	56	21.1	354	1 YFIO_ECOLI	P17443 escherichia
13	55.5	20.9	261	1 CATC_MOUSE	P28293 escherichia
14	55.5	20.9	692	1 VP3_ROTTC	P26192 porcine rot
15	55	20.6	290	1 Y290_LAMB	P03766 bacterioph
16	55	20.8	402	1 SEIP_BOVIN	P49907 bos taurus
17	55	20.8	468	1 ALS2_CANAL	O74657 candida alb
18					

19	55	20.8	828	1 YKR6_YEAST	P34239 saccharomyc
20	55	20.8	888	1 SN6A_MOUSE	O35464 mus musculu
21	55	20.8	1260	1 ALS1_CANAL	P46590 candida alb
22	55	20.8	1419	1 ALAL_CANAL	O13368 candida alb
23	54	20.2	201	1 YAHF_ECOLI	P77736 escherichia
24	54	20.4	294	1 ANPM_METJA	O58725 methanococc
25	54	20.2	296	1 MYDM_MOUSE	O35682 mus musculu
26	54	20.4	540	1 Z136_HUMAN	P52737 homo sapien
27	54	20.2	2733	1 RRPB_CYNAS	P16342 murine coro
28	53.5	20.2	187	1 UCRI_RHOSH	Q02762 rhodobacter
29	53.5	20.0	270	1 CB12_PETHY	P13869 petunia hyb
30	53.5	20.0	340	1 RIR2_MYCGE	P47471 mycoplasma
31	53.5	20.0	352	1 RADA_METJA	Q49593 methanococc
32	53.5	20.2	357	1 CARA_LACIC	O914n5 lactococcu
33	53.5	20.0	420	1 MP13_MESAU	P48968 mesocricetu
34	53.5	20.2	424	1 GCH2_CHLTR	O84736 chlamydia t
35	53.5	20.0	623	1 EXAA_PSEAE	O94437 pseudomonas
36	53	19.9	312	1 YPR1_YEAST	Q12438 saccharomyc
37	53	19.9	330	1 FBP_NEIGO	P17259 salmonella g
38	53	20.0	354	1 MURG_SALTI	O82999 salmoneila
39	53	20.0	354	1 MURG_SALTY	Q8zru3 salmonella
40	53	20.0	1119	1 ALS3_CANAL	O74623 candida alb
41	53	19.9	2731	1 RRPB_CVMJH	P29982 murine coro
42	52.5	19.8	266	1 TERM_RPH03	Q37883 bacterioph
43	52.5	19.8	327	1 COL7_ARATH	O9symb2 arabidopsis
44	52.5	19.7	339	1 RIR2_MYCPN	P75461 mycoplasma
45	52.5	19.7	354	1 PONI_MOUSE	P52430 mus musculu

## ALIGNMENTS

### RESULT 1

RRPO_SHVX	RRPO_SHVX	STANDARD;	PRT; 1718 AA.
AC	Q04575;		
DT	01-JUN-1994 (Rel. 29, Created)		
DT	01-JUN-1994 (Rel. 29, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	RNA replication protein (194 kDa protein) (ORF 1) [Contains: RNA-		
DE	directed RNA polymerase (EC 2.7.7.48); Probable helicase].		
OS	Shallot virus X (ShVX).		
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Allexivirus.		
OX	NCBI_TaxID=31770;		
RN	[1]		
RP	SEQUENCE FROM N. A.		
RX	MEDLINE-93019008; PubMed-139468;		
RA	Kanyuka E.V., Vishnichenko V.K., Levay K.E., Kondrikov D.Y.,		
RA	Ryabov E.V., Zavriv S.K.;		
RT	"Nucleotide sequence of shallot virus X RNA reveals a 5'-proximal		
RT	cistron closely related to those of potexviruses and a unique		
RT	arrangement of the 3'-proximal cistrons."		
RL	J. Gen. Virol. 73:2553-2560(1992).		
CC	-!- FUNCTION: RNA-replication. The central part of this protein		
CC	possibly functions as an ATP-binding helicase.		
CC	-!- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +		
CC	{RNA}(N).		

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to licenseseib-sib.ch).  
-----  
EMBL; M92264; AAA47787.1; -  
PIR; J01734; J01734.  
InterPro; IPR005123; 2OG-FelI\_Oxy.  
InterPro; IPR001788; RNA\_dep\_RNAPol2.  
InterPro; IPR000606; Viral\_helicase.  
Pfam; PF00978; RNA\_dep\_RNAPol2; 1.  
Pfam; PF01443; Viral\_helicase; 1.  
Pfam; PF03171; 2OG-FelI\_Oxy; 1.

KW ATP-binding; Helicase; RNA replication; RNA-directed RNA polymerase;  
KW Transferrase.  
FT NP\_BIND 915 922 ATP (POTENTIAL).  
SQ SEQUENCE 1718 AA; 194529 MW; 820FEAE1E862415C CRC64;  
Alignment Scores:  
Pred. No.: 0.426 Length: 1718  
Score: 68.00 Matches: 17  
Percent Similarity: 53.19% Conservative: 8  
Best Local Similarity: 36.17% Mismatches: 20  
Query Match: 25.66% Indels: 2  
DB: 1 Gaps: 1  
US-09-856-221-2 (1-154) x RRPQ\_SHVX (1-1718)  
QY 10 CGGATCATTAATACGATTAAGTGGCGCAACTGCTCAACGGCGGCTTGGCGGCGAAA 69  
Db 858 ArgGlyIleTyrThrHisLysIleAsp-----AsnArgArgAlaThrAlaTyrMetSer 875  
QY 70 CGPACTGCAATTCGTTACCGCTTTGCTTCCTCCAGATTAACAAACAACTGCAAGT 129  
Db 876 AspValLysAsnAsnLeuThrGlyLeuValLeuProLysLeuAspArgAspLeuLeuSer 895  
QY 130 TACTGGCAGACGTTAGCACAA 150  
Db 896 SerTrrpValAlaLeuAlaGlu 902  
RESULT 2  
FTSH\_CHLVU STANDARD; PRT; 1720 AA.  
ID FTSH\_CHLVU  
AC P56369;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Cell division protein ftsh homolog.  
GN FTSH.  
OS Chlorella vulgaris.  
OC Chloroplast.  
OC Eukaryota; Viridiplantae; Chlorophyta; Trebouxiophyceae; Chlorellales;  
OC Chlorellaceae; Chlorella.  
OX NCBI\_TaxID=3077;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-IAM C-27 / Tamiya;  
RX MEDLINE-97303241; PubMed-9159184;  
RA Wakasugi T., Nagai T., Kapoor M., Sugita M., Ito M., Ito S.,  
RA Tsudzuki J., Nakashima K., Tsudzuki T., Suzuki Y., Hamada A., Ohta T.,  
RA Inamura A., Yoshinaga K., Suglura M.;  
RT "Complete nucleotide sequence of the chloroplast genome from the  
RT green alga Chlorella vulgaris: the existence of genes possibly  
RT involved in chloroplast division."  
RL Proc. Natl. Acad. Sci. U.S.A. 94:5967-5972(1997).  
CC -1- SUBCELLULAR LOCATION: CHLOROPLAST; INTEGRAL MEMBRANE PROTEIN  
CC (POTENTIAL).  
CC -1- DOMAIN: LACKS THE ZINC PROTEASE DOMAIN OF OTHER FTSH PROTEINS.  
CC ALSO MUCH LONGER IN BOTH THE N- AND C-TERMINUS.  
CC -1- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AB001684; BAA57905.1;  
CC EMBL; AB001684; BAA57906.1; ALT INIT.  
CC InterPro; IPR003593; AAA\_ATPase.  
CC InterPro; IPR003959; AAA\_ATPase\_cent.  
CC InterPro; IPR003960; AAA\_sub.  
CC InterPro; IPR000642; Peptidase\_M41.  
CC Pfam; PF00004; AAA; 1.

Pfam; PF01434; Peptidase\_M41; 1.  
DR SMART; SMO0382; AAA; 1.  
DR PROSITE; PS00674; AAA; 1.  
KW Cell division; ATP-binding; Transmembrane; Chloroplast.  
FT TRANSMEM 48 68 POTENTIAL.  
FT TRANSMEM 896 916 POTENTIAL.  
FT TRANSMEM 973 993 POTENTIAL.  
FT TRANSMEM 1021 1041 POTENTIAL.  
SQ SEQUENCE 1720 AA; 197172 MW; 0F1EA926B799D5BB CRC64;  
Alignment Scores:  
Pred. No.: 1.76 Length: 1720  
Score: 63.50 Matches: 19  
Percent Similarity: 43.40% Conservative: 4  
Best Local Similarity: 35.85% Mismatches: 19  
Query Match: 23.78% Indels: 11  
DB: 1 Gaps: 2  
US-09-856-221-2 (1-154) x FTSH\_CHLVU (1-1720)  
QY 151 GTTGCGTAAGCTGCTGCCAATACTTGCAGTTTGTATTTATCTGAGGAGCAACAAG 92  
Db 669 ValIleLeuAspSerLysLysAsnPhe-----AspTrrpSerThrLys 682  
QY 91 CGGTTAAGCAATTTGCAG-----TACGTTTCGCGCAAGCGCGCGTT 47  
Db 683 HisGlnThrAsnLeuGlnLeuTrrpPheGlnLysTyrValSerProLeuAsnProLeuVal 702  
QY 46 GAGCAGTTGCCGATTTTATGCTGATATGATACAGAA 8  
Db 703 GlnPheGlnGlyAsnPhePheCysGluGluSerValGlu 715  
RESULT 3  
YZ64\_SYNV3  
ID YZ64\_SYNV3 STANDARD; PRT; 477 AA.  
AC P73436;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Hypothetical protein sll1464.  
GN SLL1464.  
OS Synechocystis sp. (strain PCC 6803).  
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.  
OX NCBI\_TaxID=1148;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC MEDLINE-97061201; PubMed-8905231;  
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,  
RA Miyajima N., Hirose M., Suglura M., Sasamoto S., Kimura T.,  
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K.,  
RA Okumura S., Shimizu S., Takeuchi C., Wada T., Watanabe A.,  
RA Yamada M., Yasuda M., Tabata S.;  
RT "Sequence analysis of the genome of the unicellular cyanobacterium  
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the  
RT entire genome and assignment of potential protein coding regions.";  
RL DNA Res. 3:109-136(1996).  
CC -1- SIMILARITY: BELONGS TO THE UPF0061 FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; D90906; BAA17476.1;  
CC InterPro; IPR003846; UPF0061.  
CC Pfam; PF02696; UPF0061; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 477 AA; 54041 MW; 81F6899B1A6D613C CRC64;  
Alignment Scores:



```

Pred. No.: 4.18 Length: 477
Score: 60.50 Matches: 19
Percent Similarity: 51.92% Conservative: 8
Best Local Similarity: 36.54% Mismatches: 18
Query Match: 22.83% Indels: 7
DB: 1 Gaps: 3

US-09-856-221-2 (1-154) x YZ64-STNY3 (1-477)
QY 4 ACCATTCGTGATCATATACAGATAAAATCGCGCAACTG-----CGTCAACGCGCGCC 57
   ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 412 ThrTrpArgSerIleTyGlnGlnValLeuLysGlnLeuProGluSerLeuAlaLa 431
QY 58 TTGCGCG--GCGAAACGACTACTGCGCAACTGTTGTTCTCTCTCTCAGATAAAC 114
   ::|||:: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 432 ValProThrValLeuGluGlnAlaAsnProLeuThrAspLeuLeuArgProArgIle--- 450
QY 115 AAAAACTGCAAACTTACTGCGAGCGCTTAGACAA 150
   ::|||:: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 451 -----GluThrValTrpGlnAlaIleAlaGlu 459

RESULT 4
NH55_CAEEL STANDARD; PRT; 455 AA.
AC O16962: O9GTF7; O9GTF8;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Nuclear hormone receptor family member nhr-55.
GN NHR-55 OR T01G6.7.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Jones K., Bradshaw H., Graves T.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 103-455 FROM N.A.
RA Bogan A., Maiba C.V., Yamamoto K., Cohen F., Sluder A.E.;
RT "Caenorhabditis elegans nuclear receptor sequences exhibit biophysical
compatibility with the ligand-binding domain fold."
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: ORPHAN NUCLEAR RECEPTOR.
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF022978; AAG24179.1; -
DR EMBL; AF273801; AAG15150.1; -
DR EMBL; AF273802; AAG15151.1; -
DR HSP; P20393; IAEY.
DR WormPep; T01G6.7; CEF13002.
DR InterPro; IPR000536; Hormone_rec_lig.
DR InterPro; IPR001628; Znf_C4steroid.
DR Pfam; PF00105; Zf-C4; 2.
DR PRINTS; PR00047; STROIDFINGER.
DR ProDom; PD000035; Znf_C4steroid; 1.
DR SMART; SM00430; HOL1; 1.
DR SMART; SM00399; ZnF_C4; 1.
DR PROSITE; PS00031; NUCLEAR_RECEPTOR; FALSE_NEG.
KW Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW Zinc-finger. 59 124 NUCLEAR RECEPTOR-TYPE.
FT DNA_BIND

```

```

FT ZN_FING 59 86 C4-TYPE.
FT ZN_FING 95 119 C4-TYPE.
FT DOMAIN 34 40 POLY-SER.
FT DOMAIN 43 50 POLY-SER.
SQ SEQUENCE 455 AA; 52956 MW; B56CEFA00D6B1A3 CRC64.

Alignment Scores:
Pred. No.: 5.71 Length: 455
Score: 59.50 Matches: 14
Percent Similarity: 55.88% Conservative: 5
Best Local Similarity: 41.18% Mismatches: 8
Query Match: 22.45% Indels: 7
DB: 1 Gaps: 2

US-09-856-221-2 (1-154) x NH55_CAEEL (1-455)
QY 29 AATGCGGCAACTGCTCAACCGCGCTTGC CGCGCAACGCTACTGCAAAATTCGTAA 88
   ::|||:: ||| ::||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 85 ArgCysGlySerAlaAsnAsnPhelysProCysArgArgAsnMet----- 99
QY 89 CCGCTTGTCTCTCTCTCAGA---TAAACAAAAAACTGCAAA 127
   ||| |||||:: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 100 ---AsnCysGluPheLeuLysAsnGlyTrpPheAsnCysLys 112

RESULT 5
MET3_SCHPO STANDARD; PRT; 490 AA.
AC P78937; O9USV5; O9P6S1;
DT 30-MAY-2000 (Rel. 39, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Sulfate adenylyltransferase (EC 2.7.7.4) (Sulfate adenylylate
transferase) (SAT) (ATP-sulfurylase).
GN SP8C28F2.01C OR SP8C27.08C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Kohnosu A., Niwa O., Yano M., Saitoh S., Katayama T., Nagao K.,
RA Yanagida M.;
RT "A pombe chromosome II cosmid 1228 sequence."
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Woodward J., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volkart G., Aert R., Robben J., Grymonprez B.,
RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.N., Paulsen I., Potashkin J.,

```



```

16-OCT-2001 (Rel. 40, Last annotation update)
DE Hydrogenase maturation protein hypF.
GN HYF OR SL0322.
OS Synchocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synchocystis.
OX NCBI_TaxID=1148;
RN
[1]
RC SEQUENCE FROM N.A.
RA MEDLINE=96127529; PubMed=8590279;
RX Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
RA Sugita M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synchocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RT region from map positions 64% to 92% of the genome.";
RL DNA Res. 2:153-166(1995)
CC
-!- FUNCTION: INVOLVED IN THE HYDROGENASE MATURATION PROCESS (BY
CC SIMILARITY).
CC
-!- SIMILARITY: BELONGS TO THE HYPF FAMILY.
CC
-!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE
CC ACYLPHOSPHATASE FAMILY.
CC
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
-----
DR EMBL; D64000; BAA10154.1; -.
DR HSP: P00818; LAPS.
DR InterPro; IPR001792; Acylphosphatase.
DR InterPro; IPR004421; HypF.
DR Pfam; PF00708; Acylphosphatase; 1.
DR Pfam; PF01300; Sua5_ycio_yrdC; 1.
DR ProDom; PD001884; Acylphosphatase; 1.
DR ProDom; PD002209; Sua5_ycio_yrdC; 1.
DR TIGRFAMs; TIGR00143; hypF; 1.
DR PROSITE; PS00150; ACYLPHOSPHATASE_1; 1.
KW Zinc-finger; Complete proteome.
FT DOMAIN 1 89 ACYLPHOSPHATASE-LIKE.
FT ZN_FING 107 131 CA-TYPE (POTENTIAL).
FT ZN_FING 157 182 CA-TYPE (POTENTIAL).
SQ SEQUENCE 767 AA; 85358 MW; FA9BA174172AB60D CRC64;

Alignment Scores:
Pred. No.: 9.45 Length: 767
Score: 58.00 Matches: 14
Percent Similarity: 64.29% Conservative: 4
Best Local Similarity: 50.00% Mismatches: 8
Query Match: 21.89% Indels: 2
DB: 1 Gaps: 1

US-09-856-221-2 (1-154) x HYPF_SYNY3 (1-767)
QY 66 GAACGTACTGCAATCGTTAAACGCGTTTGTTCCTCCTCAGATAACAAACAACTGCA 125
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 81 GluSerPheThrAspPheThrIleArg-----ProSerSerAspGlyProLysThrAla 98

QY 126 AAGTTACTGGCAGCGTACCA 149
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 99 SerIleLeuProAspLeuSerThr 106

RESULT 8
ID Y07A_BPT4
AC P39226; STANDARD; PRT; 87 AA.
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical 10.2 kDa protein in e-segB intergenic region.
GN Y07A OR E.8 OR MSPI.

13-
; OC Bacteriophage T4.
; OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;
; OC T4-like viruses.
; OX NCBI_TaxID=10665;
; RN [1]
; RP SEQUENCE FROM N.A.
; RC STRAIN=D;
; RA Anderson B., Zurabishvili T., Marusich E., Schneider M., Mullins T.,
; RA Napuli A., Mesyanzhinov V.V., Neitzel J., Kutter E.;
; RT "Analysis of the region between lysozyme and the tRNA genes of
; RT bacteriophage T4.";
; RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.
; RN [2]
; RP SEQUENCE FROM N.A.
; RA Kutter E., Arisaka F., Kunisawa T., Tsugita A., Mosig G.,
; RA Mesyanzhinov V., Ruger W., Stidham T., Thomas E.;
; RT "Bacteriophage T4 genome analysis.";
; RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
; CC
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
-----
DR EMBL; L13089; AAB59286.1; -.
DR EMBL; AF158101; AAD42576.1; -.
KW Hypothetical protein.
SQ SEQUENCE 87 AA; 10199 MW; 2E8969F2D2D5E4B47 CRC64;

Alignment Scores:
Pred. No.: 9.68 Length: 87
Score: 57.50 Matches: 10
Percent Similarity: 53.49% Conservative: 13
Best Local Similarity: 23.26% Mismatches: 19
Query Match: 21.54% Indels: 1
DB: 1 Gaps: 1

US-09-856-221-2 (1-154) x Y07A_BPT4 (1-87)
QY 130 AACTTTCAGTTTTTTGTTTATCTGAGGAGAACAAAGCGTTTACGAATTGCAGTAC 71
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 9 AsnPhePheMetLysCysValSerGluAspPheGlyArgThrValAsnAspIleLysVal 28

QY 70 GTTTCGCGCAAGCGCGCGTTCGACGCGTTCGACGCGTTCGTCGTCGTCGTCGTCGTCG 11
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 29 IleGlyProAspHisProMetPheGluThrTyrAlaValMet---GlyAsnGluAspGly 47

QY 10 GAATGGGTTT 2
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 48 GlnIrrplyr 50

RESULT 9
MYDML_HUMAN
ID MYDML_HUMAN STANDARD; PRT; 322 AA.
AC Q96S97; Q96CS6;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Myeloid-associated differentiation marker (SBI35).
GN MYADM.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Li N., Wan T., Zhang W., Cao X.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.

```

RC TISSUE=Brain;  
RA Strausberg R.;  
RL Submitted (Sep-2001) to the EMBL/GenBank/DDHJ databases.  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).  
CC -!- SIMILARITY: BELONGS TO THE MAL FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: AY037147; AAK67628.2; -  
CC EMBL: BC013995; AAL13995.1; -  
CC TRANSMEMBRANE.  
KW  
FT TRANSMEM 41 61 POTENTIAL.  
FT TRANSMEM 70 90 POTENTIAL.  
FT TRANSMEM 101 121 POTENTIAL.  
FT TRANSMEM 137 157 POTENTIAL.  
FT TRANSMEM 171 191 POTENTIAL.  
FT TRANSMEM 203 223 POTENTIAL.  
FT TRANSMEM 239 259 POTENTIAL.  
FT TRANSMEM 294 314 POTENTIAL.  
SQ SEQUENCE 322 AA; 35273 MW; A92142CC63625A4E CRC64;  
  
Alignment Scores:  
Pred. No.: 12.3 Length: 322  
Score: 57.00 Matches: 15  
Percent Similarity: 55.00% Conservative: 7  
Best Local Similarity: 37.50% Mismatches: 18  
Query Match: 21.35% Indels: 0  
DB: 1 Gaps: 0  
  
US-09-856-221-2 (1-154) x MYDM\_HUMAN (1-322)  
QY 145 CTAACGTCGCCAGTACTTGGAGTTTGTGTTATCTGAGGAGGACAAACGCGTTA 86  
Db 103 IieThrPheAlaCysTyrrAlaAlaLeuPheCysLeuSerAlaSerIleIleTyrrProThr 122  
QY 85 ACGAATTGTCAGTACGTTGCGCGCAAGCGCGGTTCGAGCGAGTTGCGCGATTAT 26  
Db 123 ThrTyrrValGlnPheLeuSerHisGlyArgSerArgAspHisAlaIleAlaThrPhe 142  
  
RESULT 10  
EPBL\_CHICK STANDARD; PRT; 984 AA.  
AC Q07494;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DE 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Ephrin type-B receptor 1 (EC 2.7.1.112) (Tyrosine-protein kinase  
receptor EPH-2) (Tyrosine kinase CEK6 receptor) (Fragment).  
GN EPBL OR CEK6.  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
OC Gallus.  
OX NCBI\_TaxID=9031;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=93288394; PubMed=8510926;  
RA Sajjadi F.G., Pasquale E.B.;  
RT \*Five novel avian Eph-related tyrosine kinases are differentially  
expressed.\*;  
RL Oncogene 8:1807-1813(1993).  
CC -!- FUNCTION: RECEPTOR FOR MEMBERS OF THE EPHRIN-B FAMILY. BINDS TO  
CC EPHRIN-BL, -B2 AND -B3. MAY BE INVOLVED IN CELL-CELL INTERACTIONS  
CC IN THE NERVOUS SYSTEM.  
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
tyrosine phosphate.

CC -!- SUBCELLULAR LOCATION: Type I membrane protein.  
CC -!- TISSUE SPECIFICITY: EXPRESSED AT HIGH LEVELS IN THE 10-DAY EMBRYO,  
CC AND IN ADULT BRAIN, LUNG, HEART AND SKELETAL MUSCLE. LOW LEVELS OF  
CC EXPRESSION DETECTED IN ALL OTHER ADULT TISSUES TESTED.  
CC -!- SIMILARITY: CONTAINS 1 SAM DOMAIN.  
CC -!- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.  
CC -!- SIMILARITY: BELONGS TO THE TYR FAMILY OF PROTEIN KINASES. EPHRIN  
RECEPTOR SUBFAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: Z19110; BAA79526.1; -  
CC HSP; P29323; IBAF.  
DR InterPro: IPR000561; EGF-like.  
DR InterPro: IPR001090; Ephrin\_receptor.  
DR InterPro: IPR000719; Euk\_pkinase.  
DR InterPro: IPR003961; FN\_III.  
DR InterPro: IPR003962; FNIII\_repeat.  
DR InterPro: IPR001660; SAM.  
DR InterPro: IPR001245; Tyr\_pkinase.  
DR InterPro: IPR001426; YKase\_receptor.  
DR Pfam: PF00041; fn3; 2.  
DR Pfam: PF00069; pkinase; 1.  
DR Pfam: PF00536; SAM; 1.  
DR Pfam: PF01404; EPH\_III; 1.  
DR PRINTS: PR00014; FNTYPEIII.  
DR PRINTS: PR00109; TYRKINASE.  
DR ProDom: PD000001; Euk\_pkinase; 1.  
DR ProDom: PD001495; Ephrin\_receptor; 1.  
DR SMART: SM00060; FN3; 2.  
DR SMART: SM00454; SAM; 1.  
DR SMART: SM00219; TYRK; 1.  
DR PROSITE: PS01186; EGF\_2; UNKNOWN\_1.  
DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM; 1.  
DR PROSITE: PS00109; PROTEIN\_KINASE\_TYR; 1.  
DR PROSITE: PS00790; RECEPTOR\_TYR\_KIN\_V\_1; PARTIAL.  
DR PROSITE: PS00791; RECEPTOR\_TYR\_KIN\_V\_2; 1.  
DR PROSITE: PS50105; SAM\_DOMAIN; 1.  
KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;  
KW Receptor; Transmembrane; Glycoprotein; Repeat.  
FT NON\_TER 1 1  
FT DOMAIN <1 541 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 542 562 POTENTIAL.  
FT DOMAIN 563 984 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 164 300 CYS-RICH.  
FT DOMAIN 301 410 FIBRONECTIN TYPE-III 1.  
FT DOMAIN 411 527 FIBRONECTIN TYPE-III 2.  
FT DOMAIN 619 882 PROTEIN KINASE.  
FT DOMAIN 911 975 SAM.  
FT SITE 982 984 PDZ-BINDING MOTIF (POTENTIAL).  
FT NP\_BIND 625 633 ATP (BY SIMILARITY).  
FT BINDING 651 651 ATP (BY SIMILARITY).  
FT ACT\_SITE 744 744 BY SIMILARITY.  
FT MOD\_RES 594 594 PHOSPHORYLATION (AUTO-) (POTENTIAL).  
FT MOD\_RES 600 600 PHOSPHORYLATION (AUTO-) (POTENTIAL).  
FT MOD\_RES 778 778 PHOSPHORYLATION (AUTO-) (POTENTIAL).  
FT MOD\_RES 928 928 PHOSPHORYLATION (AUTO-) (POTENTIAL).  
SQ SEQUENCE 984 AA; 109519 MW; EF06C83BB63A13A1 CRC64;  
  
Alignment Scores:  
Pred. No.: 15.4 Length: 984  
Score: 56.50 Matches: 17  
Percent Similarity: 41.51% Conservative: 5  
Best Local Similarity: 32.08% Mismatches: 27  
Query Match: 21.16% Indels: 4  
DB: 1 Gaps: 1





DR	EMBL; X78544; CAA55290.1; -.
DR	PIR; S23170; S23170.
DR	HSSP; P08311; ICGH.
DR	MEROFS; S01.133; -.
DR	MGD; MGI:88563; Ctsg.
DR	InterPro; IPRO01314; Chymotrypsin.
DR	InterPro; IPRO01254; Ser.protease_Try.
DR	Pfam; PF00089; trypsin; 1.
DR	PRINTS; PR00722; CHYMOTRYPSIN.
DR	SMART; SM00020; Tryp_SPC; 1.
DR	PROSITE; PS02440; TRYPsin_HIS; 1.
DR	PROSITE; PS00134; TRYPsin_DOM; 1.
DR	PROSITE; PS00135; TRYPsin_SER; 1.
KW	Hydrolase; Serine protease; zymogen; Glycoprotein; Signal;
KW	Intermediate filament; Membrane.
FT	SIGNAL 1 18 BY SIMILARITY.
FT	PROPEP 19 20 ACTIVATION PEPTIDE.
FT	CHAIN 21 261 CATHEPSIN G.
FT	ACT_SITE 64 64 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE 108 108 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE 201 201 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	DISULFID 49 65 BY SIMILARITY.
FT	DISULFID 142 207 BY SIMILARITY.
FT	DISULFID 172 186 BY SIMILARITY.
FT	CARBOHYD 71 71 N-LINKED GLCNAC... (POTENTIAL).
FT	CONFLICT 51 51 G -> S (IN REF. 4).
FT	CONFLICT 56 56 E -> G (IN REF. 4).
FT	CONFLICT 60 60 L -> P (IN REF. 4).
SQ	SEQUENCE 261 AA; 29095 MW; 5EFA1A6E10E1DFC CRC64;

Alignment Scores:

Pred. No.:	19.4	Length:	261
Score:	55.50	Matches:	17
Percent Similarity:	57.14%	Conservative:	7
Best Local Similarity:	40.48%	Mismatches:	11
Query Match:	20.94%	Indels:	7
DB:	1	Gaps:	2

US-09-856-221-2 (1-154) x CATG\_MOUSE (1-261)

QY	1	CAAACATTCGTGGATCATATCACGATAAATGCAGCAATCGCCAACTGGTCAACGCCGCGCCTTG	60
Db	103	GLnAsnIleArgAsnAspIleMet-----LeuLeuGlnLeuArgArgArgAla-----	118
QY	61	CGCGCGAAGCTAGCTGCAAAATCGTTAACCGCTTTGTTCCTCTCCAGATAACAAAAA	120
Db	119	-----ArgArgSerGlySerValLysProValAlaLeuProGlnAlaSerLysLys	135
QY	121	CTGCAA 126	
Db	136	LeuGln 137	

RESULT 15

ID	VP3_ROTPC	STANDARD;	PRT;	692 AA.
AC	P26192;			
DT	01-MAY-1992 (rel. 22, Created)			
DT	01-MAY-1992 (rel. 22, Last sequence update)			
DE	Inner core protein VP3.			
GN	S4.			
OS	Porcine rotavirus (group C / strain Cowden).			
OC	Viruses; dsRNA viruses; Reoviridae; Rotavirus.			
OX	NCBI_TaxID=10916;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=92124743; PubMed=1310192;			
RA	Bremont M., Juste-Lesage P., Chabanne-Vautherot D.,			
RA	Charpienne A., Cohen J.;			
RT	"Sequences of the four larger proteins of a porcine group C rotavirus			
RT	and comparison with the equivalent group A rotavirus proteins.*;"			
RL	Virology 186:684-692(1992).			
RC	-!- FUNCTION: BINDS GTP AND COULD PLAY A ROLE IN THE CAPING			

```

CC MECHANISM (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: INNER CORE.
CC -1- SIMILARITY: TO OTHER VP3 PROTEINS.
CC -1- SIMILARITY: SIGNIFICANT, TO THE PROKARYOTIC ADENINE-SPECIFIC
CC METHYLTRANSFERASE FOKI PROTEIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M742119; AAA99239.1; -.
CC DR PIR; C40822; P3XRPC.
CC KW CTP-binding.
CC SQ SEQUENCE 692 AA; 81378 MW; D679352CF22FE7DE CRC64;

Alignment Scores:
Pred. No.: 20.6 Length: 692
Score: 55.50 Matches: 16
Percent Similarity: 43.64% Conservative: 8
Best Local Similarity: 29.09% Mismatches: 16
Query Match: 20.94% Indels: 15
DB: 1 Gaps: 2

US-09-856-221-2 (1-154) x VP3_ROTFC (1-692)
Qy 16 CATTATCAGCATATAAATCGCGCAACTCGGTCAACGCGCGCTTCGCGGCGAAACGTACT 75
Db 183 AsnPheValHisPheMetArgLeuLeuArgMetArgPheAlaValPropheAspGlnLeu 202
Qy 76 GCAATTCGTAAAC-----GCTTGTTC----- 99
Db 203 SerAsnArgValThrArgSerArgAlaPhePheLysSerLysIleHisIleGlyLeuArg 222
Qy 100 -----CTTCCTCAGATAAACAACAACTGCAAGTACTGG 135
Db 223 AsnGluSerIleProGlnAlaLeuAspAsnIleAsnSerGlnTrp 237

```

Search completed: January 15, 2003, 15:46:27  
Job time : 6.64274 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: January 15, 2003, 15:44:13 ; Search time 14.5082 Seconds  
(without alignments)  
4374.264 Million cell updates/sec

Title: us-09-856-221-2

Perfect score: 265

Sequence: 1 caaacattcgtgacatta.....gcagcgttagcacacgcc 154

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame\_n2p.model -DEV=xlp  
-O=/cgn2.1/USTO.spool/US09856221/runat\_15012003\_153925\_12457/app\_query.fasta\_1.1948  
-DB=SPTRMBL\_21\_QFMT=fastan -SUFFIX=rspt -MINMATCH=0.1 -LOOPCL=0 -LCOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09856221.ecgn\_1.1.171\_erunat\_15012003\_153925\_12457 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=10 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phage.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_rvirus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
1	233	87.9	2523	2	Q93RPO		Q93rpo xenorhabdus

2	137	51.7	2376	2	Q9F923		Q9f923 serratia en
3	126.5	47.7	2504	2	O85160		O85160 photorhabdu
4	118.5	44.7	2516	2	Q9RN43		Q9rn43 photorhabdu
5	102	38.5	2538	2	Q93RN7		Q93rn7 xenorhabdus
6	67	25.3	1625	12	O55597		O55597 garlic viru
7	66.5	24.9	338	16	O8U8D8		O8u8d8 aorobacteri
8	65	24.5	268	16	O9HW74		O9hw74 pseudomonas
9	63.5	24.0	886	5	Q9V6C3		Q9vc3 drosophila
10	63	23.8	1550	12	O55601		O55601 garlic viru
11	62	23.4	275	16	O8YQD5		O8yqd5 anabaena sp
12	62	23.2	339	2	Q9R2H9		Q9r2h9 pseudomonas
13	61	22.8	108	12	O9DH37		O9dh37 turkey herp
14	61	23.0	1557	12	O8QXW5		O8qxw5 garlic viru
15	60.5	22.8	269	3	Q9P696		Q9p696 neurospora
16	60.5	22.8	658	5	Q9NH88		Q9nh88 drosophila
17	60.5	22.8	658	5	Q9NFR7		Q9nfr7 drosophila
18	60.5	22.8	658	5	Q9VB22		Q9vb22 drosophila
19	59.5	22.3	227	10	O8R2B3		O8r2b3 oryza sativ
20	59.5	22.3	652	5	O96418		O96418 riptortus c
21	59.5	22.5	901	16	O9KE04		O9ke04 bacillus ba
22	59	22.3	250	5	Q9NM45		Q9nm45 leishmania
23	59	22.3	278	16	O8YPN8		O8ypn8 anabaena sp
24	59	22.1	335	16	O83180		O83180 treponema p
25	58.5	22.1	194	5	P91787		P91787 drosophila
26	58.5	21.9	317	16	O98DT3		O98dt3 rhizobium 1
27	58.5	21.9	585	17	O8TH08		O8th08 pyrococcus
28	58.5	21.9	589	16	O9A357		O9a357 caulobacter
29	58.5	21.9	609	10	O9CAFI		O9cafi arabidopsis
30	58.5	22.1	657	16	O9JUU5		O9juu5 neisseria m
31	58.5	21.9	676	5	O9W1Q6		O9w1q6 drosophila
32	58.5	21.9	840	10	O9LPP2		O9lpp2 arabidopsis
33	58	21.7	205	16	O9KPI1		O9kpi1 vibrio chol
34	58	21.7	347	17	O59074		O59074 pyrococcus
35	58	21.9	573	5	Q19563		Q19563 caenorhabdi
36	58	21.9	818	5	Q26059		Q26059 pacifastacu
37	58	21.9	886	16	O8X9F6		O8x9f6 escherichia
38	58	21.9	5157	3	Q01135		Q01135 metarhizium
39	57.5	21.7	661	16	O9JZF3		O9jzf3 neisseria m
40	57	21.3	307	16	O06723		O06723 bacillus su
41	57	21.5	611	10	O8S5U2		O8s5u2 oryza sativ
42	57	21.5	619	10	O9CAA5		O9caa5 arabidopsis
43	57	21.3	894	2	O8VME8		O8vme8 pseudomonas
44	56.5	21.2	163	10	O8S0H1		O8s0h1 oryza sativ
45	56.5	21.2	177	16	Q9ABB4		Q9abb4 caulobacter

#### ALIGNMENTS

RESULT 1  
Q93RPO ID Q93RPO PRELIMINARY; PRT; 2523 AA.  
AC Q93RPO;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
DE Xptal protein.  
GN XPTAL.  
OS Xenorhabdus nematophilus.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Xenorhabdus.  
OX NCBI\_TaxID=628;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=PMF1296;  
RX MEDLINE=21218513; PubMed=11319082;  
RA Morgan J.A.W., Sergeant M., Ellis D., Ousley M., Jarrett P.;  
RT "Sequence analysis of insecticidal genes from Xenorhabdus nematophilus  
PMF1296.";  
RL Appl. Environ. Microbiol. 67:2062-2069 (2001).  
DR EMBL: AJ308438; CAC38401.1; -;  
DR InterPro: IPR000953; Chromo.  
DR SMART: SM00298; CHROMO.1  
SQ SEQUENCE 2523 AA; 286999 MW; 3159852E0655B5B1 CRC64;

Alignment Scores:  
 Pred. No.: 2,77e-24 Length: 2523  
 Score: 233.00 Matches: 45  
 Percent Similarity: 90.20% Conservative: 1  
 Best Local Similarity: 88.24% Mismatches: 5  
 Query Match: 87.92% Indels: 0  
 DB: 2 Gaps: 0

US-09-856-221-2 (1-154) x Q93RP0 (1-2523)

QY 1 CAACACATTCGTGATCATTTATCAGCATATAAATGCGGCACTGCTCAACGCGGCGCTTG 60  
 |||||:||||| |||||:||||| |||||:||||| |||||:||||| |||||:||||| |||||:|||||  
 Db 1932 GlnThrMetArgHisLysTyrglnHisLysMetLeuGlnLeuArgGlnArgAlaLeu 1951  
 |||||:||||| |||||:||||| |||||:||||| |||||:||||| |||||:||||| |||||:|||||  
 QY 61 CCGGCGAAGCTACTGCAAAATTCGTTAAGCGCTTTGTTCTCTCAGATAAACAACAAA 120  
 |||||:||||| |||||:||||| |||||:||||| |||||:||||| |||||:||||| |||||:|||||  
 Db 1952 ProThrLysArgThrAlaAsnSerLeuThrAlaLeuPheLeuProGlnIleAsnLysLys 1971  
 |||||:||||| |||||:||||| |||||:||||| |||||:||||| |||||:||||| |||||:|||||  
 QY 121 CTGCAAGTTACTGGCAGAGCTTAGCACACGC 153  
 |||||:||||| |||||:||||| |||||:||||| |||||:||||| |||||:||||| |||||:|||||  
 Db 1972 LeuGlnGlyTyrTrpGlnThrLeuThrGlnArg 1982  
 |||||:||||| |||||:||||| |||||:||||| |||||:||||| |||||:||||| |||||:|||||

RESULT 2

ID Q9F9Z3 PRELIMINARY; PRT; 2376 AA.  
 AC Q9F9Z3;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE Sepa.  
 GN SEPA.  
 OS Serratia entomophila.  
 OG Plasmid pADAP.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Serratia.  
 OX NCBI\_TaxID=42906;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ALM02;  
 RX MEDLINE=20416224; PubMed=10960097;  
 RA Hurst M.R., Glare T.R., Jackson T.A., Ronson C.W.;  
 RT "Plasmid-located pathogenicity determinants of Serratia entomophila,  
 the causal agent of amber disease of grass grub, show similarity to  
 the insecticidal toxins of Photobacterium luminescens.";  
 RL J. Bacteriol. 182:5127-5138(2000).  
 DR EMBL; AF135182; AAG09642.1; -;  
 KW Plasmid.  
 SQ SEQUENCE 2376 AA; 262649 MW; E418DACE22DBB7BF CRC64;

Alignment Scores:  
 Pred. No.: 1.73e-10 Length: 2376  
 Score: 137.00 Matches: 28  
 Percent Similarity: 66.67% Conservative: 6  
 Best Local Similarity: 54.90% Mismatches: 17  
 Query Match: 51.70% Indels: 0  
 DB: 2 Gaps: 0

US-09-856-221-2 (1-154) x Q9F9Z3 (1-2376)

QY 1 CAACACATTCGTGATCATTTATCAGCATATAAATGCGGCACTGCTCAACGCGGCGCTTG 60  
 |||||:||||| |||||:||||| |||||:||||| |||||:||||| |||||:||||| |||||:|||||  
 Db 1785 GluValThrArgArgPtyrGlnGluAlaLeuAlaValAlaArgGlnValProAla 1804  
 |||||:||||| |||||:||||| |||||:||||| |||||:||||| |||||:||||| |||||:|||||  
 QY 61 CCGGCGAAGCTACTGCAAAATTCGTTAAGCGCTTTGTTCTCTCAGATAAACAACAAA 120  
 |||||:||||| |||||:||||| |||||:||||| |||||:||||| |||||:||||| |||||:|||||  
 Db 1805 ProGluThrArgThrAlaAsnSerLeuThrAlaLeuPheLeuProGlnGlnAsnGluVal 1824  
 |||||:||||| |||||:||||| |||||:||||| |||||:||||| |||||:||||| |||||:|||||  
 QY 121 CTGCAAGTTACTGGCAGAGCTTAGCACACGC 153  
 |||||:||||| |||||:||||| |||||:||||| |||||:||||| |||||:||||| |||||:|||||  
 Db 1825 LeuLysGlyTyrTrpGlnThrLeuAlaGlnArg 1835  
 |||||:||||| |||||:||||| |||||:||||| |||||:||||| |||||:||||| |||||:|||||

RESULT 3

085160  
 ID 085160 PRELIMINARY; PRT; 2504 AA.  
 AC 085160;  
 DT 01-NOV-1998 (TREMBLrel. 08, Created)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE Insecticidal toxin complex protein TcBa.  
 GN TCBA.  
 OS Photobacterium luminescens (Xenorhabdus luminescens).  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Photobacterium.  
 OX NCBI\_TaxID=29488;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-W-14;  
 RX MEDLINE=98306208; PubMed=9641921;  
 RA Bowen D., Rocheleau T.A., Blackburn M., Andreev O., Golubeva E.,  
 RA Bhartia R., ffrench-Constant R.H.;  
 RT "Insecticidal toxins from the bacterium Photobacterium luminescens.";  
 RL Science 280:2129-2132(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-W14;  
 RX MEDLINE=21185117; PubMed=11286884;  
 RA Waterfield N.R., Bowen D.J., Fetherston J.D., Perry R.D.,  
 RA ffrench-Constant R.H.;  
 RT "The tc genes of Photobacterium: a growing family.";  
 RL Trends Microbiol. 9:185-191(2001).  
 DR EMBL; AF047457; AAC38627.1; -;  
 DR EMBL; AF346498; AAL18460.1; -;  
 DR InterPro; IPR000566; Lipocin\_cytfABP.  
 DR PROSITE; PS00213; LIPOCALIN; UNKNOWN\_1.  
 SQ SEQUENCE 2504 AA; 280654 MW; 8F29B1693D1047CE CRC64;

Alignment Scores:  
 Pred. No.: 5.64e-09 Length: 2504  
 Score: 126.50 Matches: 28  
 Percent Similarity: 64.71% Conservative: 5  
 Best Local Similarity: 54.90% Mismatches: 17  
 Query Match: 47.74% Indels: 1  
 DB: 2 Gaps: 1

US-09-856-221-2 (1-154) x 085160 (1-2504)

QY 1 CAACACATTCGTGATCATTTATCAGCATATAAATGCGGCACTGCTCAACGCGGCGCTTG 60  
 |||||:||||| |||||:||||| |||||:||||| |||||:||||| |||||:||||| |||||:|||||  
 Db 1911 GlnGlnValArgGlnGlnValLeuThrGlnLeuArg---LeuAsnSerArgValLysThr 1929  
 |||||:||||| |||||:||||| |||||:||||| |||||:||||| |||||:||||| |||||:|||||  
 QY 61 CCGGCGAAGCTACTGCAAAATTCGTTAAGCGCTTTGTTCTCTCAGATAAACAACAAA 120  
 |||||:||||| |||||:||||| |||||:||||| |||||:||||| |||||:||||| |||||:|||||  
 Db 1930 ProLeuLeuGlyThrAlaAsnSerLeuThrAlaLeuPheLeuProGlnGlnAsnSerLys 1949  
 |||||:||||| |||||:||||| |||||:||||| |||||:||||| |||||:||||| |||||:|||||  
 QY 121 CTGCAAGTTACTGGCAGAGCTTAGCACACGC 153  
 |||||:||||| |||||:||||| |||||:||||| |||||:||||| |||||:||||| |||||:|||||  
 Db 1950 LeuLysGlyTyrTrpArgThrLeuAlaGlnArg 1960  
 |||||:||||| |||||:||||| |||||:||||| |||||:||||| |||||:||||| |||||:|||||

RESULT 4

ID Q9RN43 PRELIMINARY; PRT; 2516 AA.  
 AC Q9RN43;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 DE TcdA.  
 GN TcdA.  
 OS Photobacterium luminescens (Xenorhabdus luminescens).  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Photobacterium.  
 OX NCBI\_TaxID=29488;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-W-14;  
 RA Merlo D.J., Wegrich L.M., Roberts J.L., Petell J.K.;





RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei V., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulyov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.H., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spter E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector T., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissensbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.Q.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
 RT "The genome sequence of Drosophila melanogaster";  
 RL Science 287:2185-2193(2000).

RN [2]  
 RP SEQUENCE FROM N.A.

RC STRAIN-BERKELEY;

RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,

RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,

RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,

RA Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,

RA Yu C., Lewis S.E., Rubin G.M., Celniker S.,

RL Submitted (Oct-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AE003695; AAF54760.1; -

DR EMBL; AY058448; AAL13677.1; -

DR FLYbase; FBgn0037998; CG4848.

SQ SEQUENCE 886 AA; 99715 MW; 2DE97D584476FBC8 CRC64;

Alignment Scores:

Pred. No.: 5.36 Length: 886  
 Score: 63.50 Matches: 18  
 Percent Similarity: 52.08% Conservative: 7  
 Best Local Similarity: 37.50% Mismatches: 20  
 Query Match: 23.96% Indels: 3  
 DB: 5 Gaps: 1

US-09-856-221-2 (1-154) x Q9VGC3 (1-886)

QY 16 CATTCAGCATTAATCGCGCAACTGCTCAACGCGCGGCTTCCGCGCAACGTACT 75

DB 74 HlSValGlnAlaAsnCysArgSerLeuAsnGluGlnGlnLeuGlyPheGlnSerThr 93

QY 76 GCAATATCGTTA-----ACCGCTTGTTCTCTCCAGATAAACAACAACTGCAA 126

DB 94 AlaAsnAlaSerAlaLysAspAlaLeuLysGluArgAsnAlaGlyLysLeuGln 113

QY 127 AGTTACTGCGACGTTAGCACAA 150

DB 114 ThrTyrGlyThrMetAlaGln 121

RESULT 10

ID O55601 PRELIMINARY; PRT; 1550 AA.

AC O55601; (TREMBlrel. 06, Created)

DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)

DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)

DE Replicase.

OS Garlic virus C.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Allxivirus.

OX NCBI\_TaxID=12431;

RN [1]

RP SEQUENCE FROM N.A.

RA MEDLINE=93389442; PubMed=8376963;

RA Sumi S., Tsuneyoshi T., Furutani H.,

RT "Novel rod-shaped viruses isolated from garlic, Allium sativum,

RT possessing a unique genome organization.";

RL J. Gen. Virol. 74:1879-1885(1993).

RN [2]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=20009173; PubMed=10542028;  
 RA Sumi S., Matsumi T., Tsuneyoshi T.,  
 RT "Complete nucleotide sequences of garlic viruses A and C, members of  
 the newly ratified genus Allxivirus";  
 RL Arch. Virol. 144:1819-1826(1999).

DR EMBL; AB010302; BAA61818.1; -

DR InterPro; IPR000871; Beta\_Lactamase\_A.

DR InterPro; IPR001788; RNA\_dep\_RNAPol2.

DR InterPro; IPR000606; Viral\_helicase.

DR Pfam; PF00978; RNA\_dep\_RNAPol2; 1.

DR Pfam; PF01443; Viral\_helicase; 1.

DR PROSITE; PS00146; BETA\_LACTAMASE\_A; 1.

SQ SEQUENCE 1550 AA; 174874 MW; 0C0A75334938F4E7 CRC64;

Alignment Scores:

Pred. No.: 6.96 Length: 1550  
 Score: 63.00 Matches: 15  
 Percent Similarity: 57.45% Conservative: 12  
 Best Local Similarity: 31.91% Mismatches: 18  
 Query Match: 23.77% Indels: 2  
 DB: 12 Gaps: 1

US-09-856-221-2 (1-154) x O55601 (1-1550)

QY 10 CGTGATCATTTATCAGCATTAATAATCGCGCAACTGCTCAACGCGGCTTCCGCGCGAAA 69

DB 699 ArgAsnCysTyrArgTyrLysLeuAspAlaLysArgAlaThrAlaPheMet-----Ser 716

QY 70 CGTACTGCAATTCGTTAACCGCTTCTCTCCAGATAAACAACAACTGCAAGT 129

DB 717 AspValLysAsnAsnLeuThrGlyLeuValLeuProlLysLeuAspArgAspLeuLeuThr 736

QY 130 TACTGCGACGTTAGCACAA 150

DB 737 SerTrpValAlaLeuAlaGlu 743

RESULT 11

ID Q8YQD5 PRELIMINARY; PRT; 275 AA.

AC Q8YQD5; (TREMBlrel. 20, Created)

DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)

DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)

DE Hypothetical protein All3898.

GN All3898.

OS Anabaena sp. (strain PCC 7120).

OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.

OX NCBI\_TaxID=103690;

RN [1]

RP SEQUENCE FROM N.A.

RA MEDLINE=21595285; PubMed=11759840;

RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,

RA Watanabe A., Iriduchi M., Ishikawa A., Kawashima K., Kimura T.,

RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,

RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,

RA Yasuda M., Tabata S.,

RT "Complete genomic sequence of the filamentous nitrogen-fixing

cyanobacterium Anabaena sp. strain PCC 7120.";

RL DNA Res. 8:205-213(2001).

DR EMBL; AP003594; BAB75597.1; -

DR InterPro; IPR000073; Abhydrolase.

DR InterPro; IPR003089; AB\_hydrolase.

DR Pfam; PF00561; abhydrolase; 1.

DR PRINTS; PR00111; ABHYDROLASE.

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 275 AA; 31078 MW; 404A151E935CA7FE CRC64;

Alignment Scores:

Pred. No.: 7.19 Length: 275  
 Score: 62.00 Matches: 17



```
US-09-856-221-2 (1-154) x Q80XW5 (1-1557)
QY 10 CQTGATCATTCAGCAATAAATGCGGCAA----- 39
||||: ||| ||| ||| |||
Db 673 ArgAsnProHisAlaHisPheLeuArgGlnTyrAspAlaGlnThrThrHisGlnGln 692
QY 40 -----CTGCGTCAACGCGCG-----GCCTTGCGGCG 66
||||: ||| ||| ||| |||
Db 693 GlnPheValAspMetAlaHisGlyLeuLysArgGlyCysTyrGlyTyrLysValAspAla 712
QY 67 AAACGTACTGCA-----AAATCGTTAACCGCTTTGTCTCCT 105
||||: ||| ||| ||| |||
Db 713 LysArgAlaThrAlaPheMetSerAspValLysAsnLeuThrGlyLeuValLeuPro 732
QY 106 CAGATAACAAACAACTGCAAGTACTGCGAGACGTTAGCACAA 150
||||: ||| ||| ||| |||
Db 733 LysLeuAspArgAspLeuSerSerTyrPValAlaLeuAlaGlu 747

RESULT 15
Q9P696
ID Q9P696 PRELIMINARY; PRT; 269 AA.
AC Q9P696;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Hypothetical 29.4 kDa protein.
GN BDI.250.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RA Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,
RA Nyakatura G., Mewes H.W., Mannhaupt G.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA German Neurospora genome project;
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL355927; CAB91274.2; -.
KW Hypothetical protein.
SQ SEQUENCE 269 AA; 29370 MW; 94947D331E95B7A0 CRC64;

Alignment Scores:
Pred. No.: 11.8 Length: 269
Score: 60.50 Matches: 16
Percent Similarity: 56.82% Conservative: 9
Best Local Similarity: 36.36% Mismatches: 16
Query Match: 22.83% Indels: 3
DB: 3 Gaps: 2

US-09-856-221-2 (1-154) x Q9P696 (1-269)
QY 13 GATCATTCAGCAATAAATGCGGCAACTGCTCAACGCGGCGCTTGGCG---GCGAAA 69
||||: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 143 ASPHis-----HisGlyArgArgArgGlyGluArgAlaLysLeuProTrpLeuHis 160
QY 70 CQTGCAAAATGCTTAACCGCTTTGTCTTCTCAGATAACAAACAACTGCAAGT 129
||||: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 161 GlyThrLysSerGlyLeuGluAlaValPheLeuGluGlyLeuGluGlnHisGlnGlyAsn 180
QY 130 TACTGCGACGACG 141
||||: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 181 GluTrpGlnSer 184
```

Search completed: January 15, 2003, 15:54:25  
Job time : 19.5082 secs





GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model  
Run on: January 15, 2003, 15:39:33 ; Search time 26.94 Seconds  
(without alignments)  
2878.683 Million cell updates/sec

Title: us-09-856-221-3  
Perfect score: 507  
Sequence: 1 aatactgtgtcaacattac.....tacggcctgcattgtgc 291

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 1816940

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp  
-Q=/cgn2.1/USPTO.spool/US09856221/runat\_15012003\_153924\_12438/app\_query.fasta\_1.1948  
-DB=A.Geneseq.101002 -QFMT=fastan -SUFFIX=rag -MINMATCH=0.1 -LOOFC1=0  
-LCOEPT=0 -UNITS=bits -START=1 END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09856221.ecgn\_1.1.60.runat\_15012003\_153924\_12438 -NCPU=3  
-NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A.Geneseq.101002.\*  
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	293	57.8	2376	22	AA97694
2	220	43.4	579	18	AAW18301
3	220	43.4	579	18	AAW56574
4	220	43.4	2516	18	AAW17899
5	220	43.4	2516	19	AAW56572
6	220	43.4	2516	22	AAW72609
7	219	43.2	2522	20	AA93729
8	216	42.6	2517	22	AAW72611
9	216	42.6	2537	22	AAW72614
10	208	41.0	573	18	AAW18303
11	208	41.0	573	19	AAW56559
12	208	41.0	2504	18	AAW17871
13	208	41.0	2504	19	AAW56557
14	208	41.0	2504	22	AAW72610
15	208	41.0	2505	22	AAW72612
16	109	21.5	1565	18	AAW18305
17	109	21.5	1565	19	AAW56568
18	77	15.2	294	23	ABP30152
19	77	15.2	295	23	ABP28317
20	71	14.0	296	23	ABP28318
21	69.5	13.0	419	22	ABG17301
22	69	12.9	288	22	ABG25746
23	68	13.4	20	19	AAW56591
24	67	13.2	1031	22	AAW48266
25	67	12.5	1345	21	AAW18284
26	66	13.0	1118	22	AAW48264
27	65.5	12.2	186	23	AAW05553
28	65.5	12.9	700	22	AAW79448
29	65.5	12.9	706	22	AAW22265
30	65.5	12.9	706	22	AAW79447
31	65	12.8	299	20	AAW22568
32	64.5	12.7	1002	22	ABW59634
33	64	12.6	464	22	ABB61748
34	64	12.6	978	20	AAW17509
35	63.5	11.9	868	22	ABG22938
36	63	11.8	140	21	AAW05359
37	63	11.8	165	21	AAW05358
38	63	11.8	191	21	AAW05357
39	63	12.4	243	18	AAW55254
40	63	11.8	338	22	ABW68636
41	63	12.4	350	18	AAW55563
42	63	11.8	719	13	AAW25069
43	63	11.8	719	15	AAW49507
44	63	11.8	719	15	AAW45775
45	63	11.8	719	16	AAW74096

# ALIGNMENTS

RESULT 1  
AA97694  
ID AAY97694 standard; Protein; 2376 AA.  
XX  
AC AAY97694;  
XX  
DT 19-JUN-2001 (first entry)  
XX  
DE SepA protein encoded by Serratia insecticidal protein complex gene.  
XX  
KW Insecticidal protein complex; amber disease; insect; Coleoptera;  
XX  
KW pesticide; SepA protein.  
XX  
OS Serratia sp.  
XX  
PN WO200116305-A2.  
XX  
PD 08-MAR-2001.  
XX

PF 04-SEP-2000; 2000WO-N200174.  
 XX  
 PR 02-SEP-1999; 99NZ-0337610.  
 XX  
 PA (NZPA-) NEW ZEALAND PASTORAL AGRIC RES INST LTD.  
 XX  
 PI Glare TR, Hurst MRH, Jackson TA;  
 XX  
 XX WPI: 2001-169009/17.  
 DR N-PSDB; AAA91292.  
 XX  
 XX New nucleic acid encoding a polypeptide useful as a pesticide  
 PT especially for Coleoptera -  
 PT  
 XX  
 XX Claim 24; Page 92-99; 109pp; English.  
 PS  
 XX This sequence represents the SepA protein encoded by the Serratia  
 CC insecticidal protein complex gene of the invention. The invention relates  
 CC to a gene encoding an insecticidal protein complex or a functional  
 CC fragment, a neutral mutation, or a homologue of the complex. The  
 CC polypeptides and nucleotides of the invention are used to induce amber  
 CC diseases or like conditions in insects, especially to those from the  
 CC order comprising Coleoptera, useful as a pesticide.  
 XX  
 XX Sequence 2376 AA;  
 SQ  
 Alignment Scores:  
 Pred. No.: 4.53e-29 Length: 2376  
 Score: 293.00 Matches: 65  
 Percent Similarity: 80.21% Conservative: 12  
 Best Local Similarity: 67.71% Mismatches: 19  
 Query Match: 57.79% Indels: 1  
 DB: 22 Gaps: 0  
 US-09-856-221-3 (1-291) x AAY97694 (1-2376)  
 QY 1 AATACCTGCTCACTTACTGACGCGAGTGCAGAGCAGCTGGCAGAAATGCTGCAA 60  
 Db 1907 AsnThrLeuLeuGlyIleThrGluArgGlnAspAlaGluAlaLeuAlaLysLeuLeuGln 1926  
 QY 61 ACTCAAGGCAGTGAATAGCTTTGTCAGAGTATTAAATGACAGGC - AAGATGATTGCTGAA 119  
 Db 1927 ThrGlnGlySerGluLeuLeuArgGlnGlyLeuArgGlnGlnAspAsnValLeuGlu 1945  
 QY 120 ATTGATGCTGATGAAGTGGCGCTTAAGAAAGCGCTCATGTTGTCACATCTCTTTTGAC 179  
 Db 1947 IleAspAlaAspIleAlaLeuGluGluSerArgGlyAlaGlnMetArgPheGlu 1966,  
 QY 180 AGCTTCAGTACGCTGTATGACGAGATGTTAACTCCGGTGAACAAAGCGATGATCTG 239  
 Db 1967 ArgTyrLysValLeuTyrGluAlaAspValAsnThrGlyGluLysGlnAlaMetAspLeu 1986  
 QY 240 TATCTCTCTTCATCGGTATTGACGACCCAGCAGTACGCGCCCTGCATATG 287  
 Db 1987 TyrLeuSerSerValLeuSerAlaSerThrAlaAlaLeuPheLeu 2002  
 RESULT 2  
 AAW18301  
 ID AAW18301 standard; Protein: 579 AA.  
 XX  
 AC AAW18301;  
 XX  
 DT 29-JAN-1998 (first entry)  
 XX  
 XX Photorhabdus luminescens insect toxin TcdA111.  
 DE  
 XX Insecticide; insect; toxin; pest control; biological control;  
 KW Photorhabdus luminescens; TcdA; Southern corn rootworm;  
 KW Colorado potato beetle; Western corn rootworm; meal worm;  
 KW boll weevil; turf grub; Coleoptera; beet armyworm; black cutworm;  
 KW cabbage looper; codling moth; corn earworm; European corn borer;  
 KW tobacco hornworm; tobacco budworm; Lepidoptera; Hymenoptera;  
 KW Diptera, Dictyoptera; Acarina; Homoptera.

XX  
 OS Photorhabdus luminescens strain W-14 (ATCC 55397).  
 XX  
 PN WO9717432-A1.  
 XX  
 PD 15-MAY-1997.  
 XX  
 PF 06-NOV-1996; 96WO-US18003.  
 XX  
 PR 28-AUG-1996; 96US-0705484.  
 PR 06-NOV-1995; 95US-0007255.  
 PR 28-FEB-1996; 96US-0608423.  
 XX  
 XX (WISC) WISCONSIN ALUMNI RES FOUND.  
 PA  
 XX Blackburn MB, Bowen DJ, Ciche TA, Ensign JC, Fatig R;  
 PI French-Constant RH, Guo L, Hey TD, Merlo DJ, Orr GL;  
 PI Petrelli J, Roberts JL, Rocheleau TA, Schoonover S;  
 PI Strickland JA;  
 XX  
 DR WPI: 1997-281022/25.  
 DR N-PSDB; AAT68845.  
 XX  
 XX Photorhabdus sp. insecticidal protein toxins and DNA encoding them -  
 PT can be genetically engineered into insect larvae food and plants for  
 PT insect control  
 PT  
 XX Claim 34; Page 213-215; 276pp; English.  
 PS  
 XX This polypeptide comprises the 63.6 kDa TcdA111 insecticidal toxin  
 CC protein of Photorhabdus luminescens W-14. Its sequence was deduced  
 CC from a genomic DNA clone (AAT68845) and includes N-terminal and  
 CC tryptic peptide sequences obtained from the isolated protein.  
 CC TcdA111 is a proteolytic cleavage product of TcdA (see AAW17899).  
 CC Claimed toxins of P. luminescens (see AAW17871, AAW17884-89,  
 CC AAW17899-900, AAW18301-06) can be produced by recombinant DNA methods and  
 CC applied to, or genetically engineered into, insect larvae food and  
 CC plants for insect control. The toxins are particularly effective  
 CC against Southern corn rootworm, Colorado potato beetle, Western corn  
 CC rootworm, meal worm, boll weevil and turf grub (Coleoptera), beet  
 CC armyworm, black cutworm, cabbage looper, codling moth, corn earworm,  
 CC European corn borer, tobacco hornworm and tobacco budworm  
 CC (Lepidoptera), and are also active against insects of the orders  
 CC Hymenoptera, Diptera, Dictyoptera, Acarina and Homoptera. (All  
 CC claimed).  
 XX  
 SQ Sequence 579 AA;  
 Alignment Scores:  
 Pred. No.: 1.02e-19 Length: 579  
 Score: 220.00 Matches: 50  
 Percent Similarity: 74.19% Conservative: 19  
 Best Local Similarity: 53.76% Mismatches: 24  
 Query Match: 43.39% Indels: 1  
 DB: 18 Gaps: 0  
 US-09-856-221-3 (1-291) x AAW18301 (1-579)  
 QY 1 AATACCTTGTCTCAACATTACTGACGCGAGTGCAGAGCAGCTGGCAGAAATGCTGCAA 60  
 Db 101 SerThrLeuGlnAsnIleIleGluArgGlnAspAlaGluAlaLeuAsnAlaLeuLeuGln 120  
 QY 61 ACTCAAGGCAGTGAATAGCTTTGTCAGAGTATTAAATGACAGGC - AAGATGATTGCTGAA 119  
 Db 121 AsnGlnAlaGluLeuIleLeuThrAsnLeuSerIleGlnAspLysThrIleGluGlu 140  
 QY 120 ATTGATGCTGATGAAGTGGCGCTTAAGGAAAGCGCTCATGTCGACAAATCTGTTTGAC 179  
 Db 141 LeuAspAlaGluLysThrValLeuGluLysSerLysAlaGlyAlaGlnSerArgPheAsp 160,  
 QY 180 AGCTTCAGTACGCTGTATGACGAGATGTTAACTCCGGTGAACAAAGCGATGATCTG 239  
 Db 161 SerTyrGlyLysLeuTyrAspGluAsnIleAsnAlaGlyGluAsnGlnAlaMetThrLeu 180







CC luminescens rather than the nematodes. The toxins have activity against  
 CC Lepidopteran insects such as Cabbage Looper (*Trichoplusia ni*),  
 CC European Corn Borer (*Ostrinia nubilalis*) and Fall Armyworm  
 CC (*Spodoptera frugiperda*) and also against Coleopteran insects  
 CC (e.g., Colorado Potato Beetle, *Leptinotarsa decimlineata*). In  
 CC addition the toxins are active against strains resistant to known  
 CC insecticides. The DNA sequence can be used to generate transgenic plants  
 CC of various species that are resistant to economically important insect  
 CC pests and also for recombinant production of the toxins for use as  
 CC insecticides.

XX SQ Sequence 2522 AA;

Alignment Scores: 2,07e-19 Length: 2522  
 Pred. No.: 219.00 Matches: 50  
 Score: 75.27% Conservative: 20  
 Percent Similarity: 53.76% Mismatches: 23  
 Best Local Similarity: 43.20% Indels: 1  
 Query Match: 20 Gaps: 0  
 DB:

US-09-856-221-3 (1-291) x AAF33729 (1-2522)

QY 1 AATACCTTGCTCAACATTACTGACGGCAGGATGCGAGACACTGGCAGAAATGCTGCAG 60  
 Db 2044 SerThrLeuGlnAsnIleIleGluArgGlnAspAlaGluAlaLeuAsnAlaLeuGln 2063  
 QY 61 ACTCAAGCGAGTGAATTAGCTTTGACAGACTATTAAATGCGAGCG-AGATGATGCTGCAA 119  
 Db 2064 AsnGlnAlaGluLeuIleLeuThrAsnLeuSerIleGlnAspIysThrIleGluGlu 2083  
 QY 120 APTGATGCTGATGAAGTGGCGCTTAAGGAAAGCCGTCATGTCGACAACTCGTTTGGAC 179  
 Db 2084 LeuAspAlaGluIysThrValLeuGluLysSerIysAlaGlyAlaGlnSerArgPheAsp 2103  
 QY 180 ACCTTCATGCTGATGATGACGAAGATGTTAACTCCGCTGAAGAAACAGCGATGATCTG 239  
 Db 2104 SerIysSerIysLeuHisAspGluAsnIleAsnAlaGlyGluAsnGlnAlaMetThrLeu 2123  
 QY 240 TATCTCTCTCATCGGTATTGACACACAGCAGTACGGCC 278  
 Db 2124 ArgAlaSerAlaAlaGlyLeuThrThrAlaValGlnAla 2136

RESULT 8  
 AAB72611  
 ID AAB72611 standard; Protein; 2517 AA.

XX AC AAB72611;

XX DT 04-MAY-2001 (first entry)

XX DE Modified Photorhabdus tcda toxin.

XX KW TcdA; TcdA; insect toxin; plant; insect resistance.

XX OS Photorhabdus sp.

XX OS Synthetic.

XX PN W0200111029-A1.

XX PD 15-FEB-2001.

XX PF 11-AUG-2000; 2000WO-US22237.

XX PR 11-AUG-1999; 99US-0148356.

XX XX (DOWC ) DOW AGROSCIENCES LLC.

XX PA Petell JK, Merlo DJ, Herman RA, Roberts JL, Guo L, Schafer BW;

XX PI Sukhapinda K, Merlo AO;

XX DR WPI; 2001-191536/19.

XX N-PSDB; AAF58780.

XX Novel polynucleotide sequence encoding insect toxins, useful for  
 PT producing transgenic plants having resistance to insects, especially  
 PT corn rootworm -

XX Claim 1; Page 72-83; 106pp; English.

XX The present invention provides the protein and coding sequences of  
 CC modified versions of the Photorhabdus tcda and TcdA toxins. These are  
 CC suitable for expression in plants. The toxins are effective against  
 CC insects upon ingestion, and the sequences provided can be used to produce  
 CC transgenic plants with insect resistance. The present sequence is the  
 CC modified Photorhabdus tcda protein.

XX SQ Sequence 2517 AA;

Alignment Scores: 5,09e-19 Length: 2517  
 Pred. No.: 216.00 Matches: 50  
 Score: 74.19% Conservative: 19  
 Percent Similarity: 53.76% Mismatches: 24  
 Best Local Similarity: 42.60% Indels: 1  
 Query Match: 22 Gaps: 0  
 DB:

US-09-856-221-3 (1-291) x AAB72611 (1-2517)

QY 1 AATACCTTGCTCAACATTACTGACGGCAGGATGCGAGACACTGGCAGAAATGCTGCAG 60  
 Db 2039 SerThrLeuGlnAsnIleIleGluArgGlnAspAlaGluAlaLeuAsnAlaLeuGln 2058  
 QY 61 ACTCAAGCGAGTGAATTAGCTTTGACAGACTATTAAATGCGAGCG-AGATGATGCTGCAA 119  
 Db 2059 AsnGlnAlaGluLeuIleLeuThrAsnLeuSerIleGlnAspIysThrIleGluGlu 2078  
 QY 120 ATTGATGCTGATGAAGTGGCGCTTAAGGAAAGCCGTCATGTCGACAACTCGTTTGGAC 179  
 Db 2079 LeuAspAlaGluIysThrValLeuGluLysSerIysAlaGlyAlaGlnSerArgPheAsp 2098  
 QY 180 ACCTTCATGCTGATGATGACGAAGATGTTAACTCCGCTGAAGAAACAGCGATGATCTG 239  
 Db 2099 SerTyrGlyLysLeuTyrAspGluAsnIleAsnAlaGlyGluAsnGlnAlaMetThrLeu 2118  
 QY 240 TATCTCTCTCATCGGTATTGACACACAGCAGTACGGCC 278  
 Db 2119 ArgAlaSerAlaAlaGlyLeuThrThrAlaValGlnAla 2131

RESULT 9

AAB72614

XX ID AAB72614 standard; Protein; 2537 AA.

XX AC AAB72614;

XX DT 04-MAY-2001 (first entry)

XX DE TcdA toxin-zeln ER signal peptide fusion protein.

XX KW TcdA; TcdA; insect toxin; plant; insect resistance.

XX OS Chimeric - Photorhabdus sp.

XX OS Chimeric - zea mays.

XX PN W0200111029-A1.

XX PD 15-FEB-2001.

XX PF 11-AUG-2000; 2000WO-US22237.

XX PR 11-AUG-1999; 99US-0148356.

XX XX (DOWC ) DOW AGROSCIENCES LLC.

XX PA Petell JK, Merlo DJ, Herman RA, Roberts JL, Guo L, Schafer BW;

XX PI Sukhapinda K, Merlo AO;







CC Southern corn rootworm, Colorado potato beetle, Western corn rootworm,  
 CC meal worm, boll weevil and turf grub (Coleoptera), beet armyworm, black  
 CC cutworm, cabbage looper, codling moth, corn earworm, European corn  
 CC borer, tobacco hornworm and tobacco budworm (Lepidoptera), and  
 CC are also active against insects of the orders Hymenoptera, Diptera  
 CC Dictyoptera, Acarina and Homoptera. (All claimed).  
 XX  
 SQ Sequence 2504 AA;

Alignment Scores:  
 Pred. No.: 5,61e-18 Length: 2504  
 Score: 208.00 Matches: 46  
 Percent Similarity: 72.92% Conservative: 24  
 Best Local Similarity: 47.92% Mismatches: 26  
 Query Match: 41.03% Indels: 1  
 DB: 18 Gaps: 0

US-09-856-221-3 (1-291) x AAW17871 (1-2504)

QY 1 AATACCTGCTCAACATTAAGCGGAGGATGAGAGGATGCGAGATGCTGCGAA 60  
 Db 2032 SerSerLeuLeuGlyTyrSerGluArgGlnAspAlaGluAlaMetSerGlnLeuLeuGln 2051  
 QY 61 ACTCAAGCAGTGAATGCTTTCGAGAGTATTAAATGACGAGC-AAGATGATGCTGAA 119  
 Db 2052 ThrGlnAlaSerGluLeuIleLeuThrSerIleArgMetGlnAspAsnGlnLeuAlaGlu 2071  
 QY 120 ATTGATGCTGATGAAGTGGCGCTTAAGGAAAGCCGCTATGTCACACATCTCTTTGAC 179  
 Db 2072 LeuAspSerGluLysThrAlaLeuGlnValSerLeuAlaGlyValGlnGlnArgPheAsp 2091  
 QY 180 ACCTTCAGTACGCTGTATGACGAGATGTTAACTCCGCTGAAACAAAGCGATGATCTG 239  
 Db 2092 SerTyrSerGlnLeuTyrGluGluAsnIleAsnAlaGlyGluArgAlaLeuAlaLeu 2111  
 QY 240 TATCTCTTCTTCATGCTGATTTGACGACGAGTACGCGCCCTGCATATG 287  
 Db 2112 ArgSerGluSerAlaIleGluSerGlnGlyAlaGlnIleSerArgMet 2127

RESULT 13

AAW56557

ID AAW56557 standard; Protein; 2504 AA.

XX AC AAW56557;

XX DT 07-AUG-1998 (first entry)

XX DE Toxin TcBa, encoded by the tcBa gene from genomic region tcb.

XX KW Photorhabdus luminescens W-14; nematode; symbiotic;

XX KW Heterorhabditis; tea; tcb; tcc; tcd; insecticidal activity; toxin;

XX KW Lepidoptera; Coleoptera; Hymenoptera; Diptera; Dictyoptera; Acarina;

XX KW Homoptera; Southern corn rootworm; Colorado potato beetle;

XX KW mealworm; boll weevil; turf grub; beetle armyworm; black cutworm;

XX KW cabbage looper; codling moth; corn earworm; European corn borer;

XX KW Tobacco hornworm; budworm.

XX OS Photorhabdus luminescens.

XX PN WO9808932-A1.

XX PD 05-MAY-1997; 97WO-US07657.

XX PF 06-NOV-1996; 96WO-US18003.

XX PR 28-AUG-1996; 96US-0705484.

XX PR 06-NOV-1996; 96US-0743699.

XX PA (DOWC ) DOWELANCO.

XX PA (WISC ) WISCONSIN ALUMNI RES FOUND.

XX PI Blackburn MB, Bowen DJ, Cliche TA, Ensign JC, Fatig R;

PI French-Constant RH, Guo L, Hey TD, Merlo DJ, Orr GL;  
 PI Petell J, Roberts JL, Rocheleau TA, Schoonover S;  
 XX Strickland JA, Sukhapingda K;  
 XX WPI; 1998-179427/16.  
 DR N-PSDB; AAV29985.  
 XX

PI Isolated toxins from Photorhabdus luminescens strains - useful for  
 PI control of insect pests

XX Claim 34; Pages 163-169; 32lpp; English.

XX The present sequence represents a protein named TcBa of the bacterium  
 CC Photorhabdus luminescens (W-14). This is a symbiotic bacterium of the  
 CC nematodes of the Heterorhabditis genus. The bacterium has at least 4  
 CC distinct genomic regions, tea, tcb, tcc, and tcd. Peptide products are  
 CC produced from these regions that are associated with insecticidal  
 CC activity. The native toxins are secreted proteins. The proteins are  
 CC toxic to insects upon exposure and especially when ingested. The  
 CC nucleic acid sequence can be used to produce transgenic plants,  
 CC baculoviruses or microbial hosts for toxin production. They can be used  
 CC to control insects pests from the Lepidoptera, Coleoptera, Hymenoptera,  
 CC Diptera, Dictyoptera, Acarina or Homoptera orders, especially the  
 CC Southern or Western corn rootworm, Colorado potato beetle, mealworm,  
 CC boll weevil, turf grub, beetle armyworm, black cutworm, cabbage looper,  
 CC codling moth, corn earworm, European corn borer or tobacco hornworm  
 CC or budworm.

XX SQ Sequence 2504 AA;

Alignment Scores:

Pred. No.: 5,61e-18 Length: 2504  
 Score: 208.00 Matches: 46  
 Percent Similarity: 72.92% Conservative: 24  
 Best Local Similarity: 47.92% Mismatches: 26  
 Query Match: 41.03% Indels: 1  
 DB: 19 Gaps: 0

US-09-856-221-3 (1-291) x AAW56557 (1-2504)

QY 1 AATACCTGCTCAACATTAAGCGGAGGATGAGAGGATGCGAGATGCTGCGAA 60

Db 2032 SerSerLeuLeuGlyTyrSerGluArgGlnAspAlaGluAlaMetSerGlnLeuLeuGln 2051

QY 61 ACTCAAGCAGTGAATGCTTTCGAGAGTATTAAATGACGAGC-AAGATGATGCTGAA 119

Db 2052 ThrGlnAlaSerGluLeuIleLeuThrSerIleArgMetGlnAspAsnGlnLeuAlaGlu 2071

QY 120 ATTGATGCTGATGAAGTGGCGCTTAAGGAAAGCCGCTATGTCACACATCTCTTTGAC 179

Db 2072 LeuAspSerGluLysThrAlaLeuGlnValSerLeuAlaGlyValGlnGlnArgPheAsp 2091

QY 180 ACCTTCAGTACGCTGTATGACGAGATGTTAACTCCGCTGAAACAAAGCGATGATCTG 239

Db 2092 SerTyrSerGlnLeuTyrGluGluAsnIleAsnAlaGlyGluArgAlaLeuAlaLeu 2111

QY 240 TATCTCTTCTTCATGCTGATTTGACGACGAGTACGCGCCCTGCATATG 287

Db 2112 ArgSerGluSerAlaIleGluSerGlnGlyAlaGlnIleSerArgMet 2127

RESULT 14

AAW72610

ID AAW72610 standard; Protein; 2504 AA.

XX AC AAW72610;

XX DT 04-MAY-2001 (first entry)

XX DE Photorhabdus tcbA toxin.

XX KW TcBa; insect toxin; plant; insect resistance.

XX OS Photorhabdus sp.

XX WO200111029-A1.  
 XX Synthetic.  
 XX WO200111029-A1.  
 XX 15-FEB-2001.  
 XX 11-AUG-2000; 2000WO-US22237.  
 XX 11-AUG-1999; 99US-0148356.  
 XX (DOWC ) DOW AGROSCIENCES LLC.  
 XX Petell JK, Merlo DJ, Herman RA, Roberts JL, Guo L, Schafer BW;  
 XX Sukhapinda K, Merlo AO;  
 XX WPI; 2001-191536/19.  
 XX N-PSDB; AAF58779.  
 XX Novel polynucleotide sequence encoding insect toxins, useful for  
 XX producing transgenic plants having resistance to insects, especially  
 XX corn rootworm  
 XX Disclosure: Page 62-72; 106pp; English.  
 XX The present invention provides the protein and coding sequences of  
 XX modified versions of the Photobabidus Tcda and Tcda toxins. These are  
 XX suitable for expression in plants. The toxins are effective against  
 XX insects upon ingestion, and the sequences provided can be used to produce  
 XX transgenic plants with insect resistance. The present sequence is the  
 XX Photobabidus Tcda protein.  
 XX SQ Sequence 2504 AA;  
 Alignment Scores:  
 Pred. No.: 5,61e-18 Length: 2504  
 Score: 208.00 Matches: 46  
 Percent Similarity: 72.92% Conservative: 24  
 Best Local Similarity: 47.92% Mismatches: 26  
 Query Match: 41.03% Indels: 1  
 DB: 22 Gaps: 0  
 US-09-856-221-3 (1-291) x AAB72610 (1-2504)  
 QY 1 AATACCTGCTCAACATTACTGACGCGAGGATGACAGACGACGAGATGCTGCAA 60  
 Db 2032 SerSerLeuLeuGlyTyrSerGluArgGlnAspAlaGluAlaMetSerGlnLeuGln 2051  
 QY 61 ACTCAAGGAGTGAATTAGCTTTGACAGAGTATTAAATGACAGGC-AAGATGATGCTGAA 119  
 Db 2052 ThrGlnAlaSerGluLeuLeuThrSerIleArgMetGlnAspAsnGlnLeuAlaGlu 2071  
 QY 120 ATTGATGCTGATGAAGTGGCGCTTAAGAAAGCCGTCATGCTGACCAATCTCTTTTGAC 179  
 Db 2072 LeuAspSerGluLysThrAlaLeuGlnValSerLeuAlaGlyValGlnArgPheAsp 2091  
 QY 180 AGCTTCAGTACGCTGATGACAGAGATGTTAACTCCGGTGAACAAACAGGATGATCTG 239  
 Db 2092 SerTyrSerGlnLeuTyrGluGluAsnIleAsnAlaGlyGluGlnArgAlaLeuAlaLeu 2111  
 QY 240 TATCTCTCTCATCGGTATTGACGACGACGATGACGCGCCCTGCATATG 287  
 Db 2112 ArgSerGluSerAlaIleGluSerGlnGlyAlaGlnIleSerArgMet 2127  
 RESULT 15  
 AAB72612  
 ID AAB72612 standard; Protein; 2505 AA.  
 XX AAB72612;  
 AC  
 XX  
 XX 04-MAY-2001 (first entry)  
 DT  
 XX Modified Photobabidus tcda toxin.  
 DE  
 XX Tcda; Tcda; insect toxin; plant; insect resistance.  
 KW

XX Photobabidus sp.  
 OS Synthetic.  
 XX WO200111029-A1.  
 XX 15-FEB-2001.  
 XX 11-AUG-2000; 2000WO-US22237.  
 XX 11-AUG-1999; 99US-0148356.  
 XX (DOWC ) DOW AGROSCIENCES LLC.  
 XX Petell JK, Merlo DJ, Herman RA, Roberts JL, Guo L, Schafer BW;  
 XX Sukhapinda K, Merlo AO;  
 XX WPI; 2001-191536/19.  
 XX N-PSDB; AAF58781.  
 XX Novel polynucleotide sequence encoding insect toxins, useful for  
 XX producing transgenic plants having resistance to insects, especially  
 XX corn rootworm  
 XX Claim 1; Page 83-93; 106pp; English.  
 XX The present invention provides the protein and coding sequences of  
 XX modified versions of the Photobabidus Tcda and Tcda toxins. These are  
 XX suitable for expression in plants. The toxins are effective against  
 XX insects upon ingestion, and the sequences provided can be used to produce  
 XX transgenic plants with insect resistance. The present sequence is the  
 XX modified Photobabidus Tcda protein.  
 XX SQ Sequence 2505 AA;  
 Alignment Scores:  
 Pred. No.: 5,61e-18 Length: 2505  
 Score: 208.00 Matches: 46  
 Percent Similarity: 72.92% Conservative: 24  
 Best Local Similarity: 47.92% Mismatches: 26  
 Query Match: 41.03% Indels: 1  
 DB: 22 Gaps: 0  
 US-09-856-221-3 (1-291) x AAB72612 (1-2505)  
 QY 1 AATACCTGCTCAACATTACTGACGCGAGGATGACAGACGACGAGATGCTGCAA 60  
 Db 2032 SerSerLeuLeuGlyTyrSerGluArgGlnAspAlaGluAlaMetSerGlnLeuGln 2052  
 QY 61 ACTCAAGGAGTGAATTAGCTTTGACAGAGTATTAAATGACAGGC-AAGATGATGCTGAA 119  
 Db 2053 ThrGlnAlaSerGluLeuLeuThrSerIleArgMetGlnAspAsnGlnLeuAlaGlu 2072  
 QY 120 ATTGATGCTGATGAAGTGGCGCTTAAGAAAGCCGTCATGCTGACCAATCTCTTTTGAC 179  
 Db 2073 LeuAspSerGluLysThrAlaLeuGlnValSerLeuAlaGlyValGlnArgPheAsp 2092  
 QY 180 AGCTTCAGTACGCTGATGACGAGATGTTAACTCCGGTGAACAAACAGGATGATCTG 239  
 Db 2093 SerTyrSerGlnLeuTyrGluGluAsnIleAsnAlaGlyGluGlnArgAlaLeuAlaLeu 2112  
 QY 240 TATCTCTCTCATCGGTATTGACGACGACGATGACGCGCCCTGCATATG 287  
 Db 2113 ArgSerGluSerAlaIleGluSerGlnGlyAlaGlnIleSerArgMet 2128  
 Search completed: January 15, 2003, 15:50:37  
 Job time : 35.94 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: January 15, 2003, 15:45:23 ; Search time 9.37561 Seconds  
(without alignments)  
1826.456 Million cell updates/sec

Title: us-09-856-221-3

Perfect score: 507

Sequence: 1 aatacctgttcacattac.....tacgccctgatgtgac 291

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Xgapop 10.0 , Xgapext 0.5  
Xgapop 6.0 , Xgapext 7.0  
Xgapop 6.0 , Xgapext 7.0

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-O=/cgn2.1/USPTO.spool/US09856221/runat\_15012003\_153926\_12521/app\_query.fasta\_1.1948  
-DB-Issued\_Patents\_AA\_QFMT-fastan -SUFFIX-rai -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=Dlcs -START=1 -END=1 -MATRIX=BIOSUM62 -TRANS=human40.cdi  
-LIST=45 -LOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09856221@cgn2.1\_24\_erunat\_15012003\_153926\_12521 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-XGAPOP=10 -XGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: /cgn2.5/ptodata/1/1aa/5A.COMB.pep.\*
- 2: /cgn2.5/ptodata/1/1aa/5B.COMB.pep.\*
- 3: /cgn2.5/ptodata/1/1aa/6A.COMB.pep.\*
- 4: /cgn2.5/ptodata/1/1aa/6B.COMB.pep.\*
- 5: /cgn2.5/ptodata/1/1aa/PCTUS.COMB.pep.\*
- 6: /cgn2.5/ptodata/1/1aa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	219	43.2	2522	4	US-09-251-645-13 Sequence 13, Appl
2	65	12.8	299	4	US-09-222-938A-34 Sequence 34, Appl
3	64	12.6	978	2	US-08-415-593-43 Sequence 43, Appl
4	63	11.8	719	1	US-07-943-843-4 Sequence 4, Appl
5	63	11.8	719	1	US-07-943-843-4 Sequence 4, Appl
6	61.5	11.5	799	4	US-09-165-396-4 Sequence 4, Appl
7	60.5	11.3	152	6	5187153-4 Patent No. 5187153
8	59	11.6	100	3	US-09-034-916-13 Sequence 13, Appl
9	58.5	10.9	169	1	US-08-741-406-2 Sequence 2, Appl
10	58.5	10.9	169	4	US-09-024-472-2 Sequence 2, Appl
11	58.5	10.9	782	4	US-09-543-084A-32 Sequence 32, Appl
12	58.5	10.9	782	4	US-09-543-084A-33 Sequence 33, Appl

c 13	58.5	10.9	782	4	US-09-543-084A-34	Sequence 34, Appl
c 14	58.5	10.9	782	4	US-09-543-084A-35	Sequence 35, Appl
c 15	58.5	10.9	782	4	US-09-543-084A-36	Sequence 36, Appl
c 16	58	11.4	2137	4	US-09-134-001C-4463	Sequence 4463, Ap
c 17	57.5	10.7	218	5	PCT-US94-14106-57	Sequence 57, Appl
c 18	57.5	10.7	218	5	PCT-US94-14106-61	Sequence 61, Appl
c 19	57	11.2	147	4	US-09-347-833-10	Sequence 10, Appl
c 20	57	10.7	366	3	US-08-875-811-55	Sequence 55, Appl
c 21	57	10.7	750	4	US-09-185-160-14	Sequence 14, Appl
c 22	56.5	11.1	1164	4	US-09-457-708-2	Sequence 2, Appl
c 23	56	10.5	75	1	US-08-289-458-2	Sequence 2, Appl
c 24	56	10.5	75	2	US-08-761-549-2	Sequence 2, Appl
c 25	56	10.5	75	4	US-09-127-646-2	Sequence 2, Appl
c 26	56	11.0	244	4	US-09-134-001C-4784	Sequence 4784, Ap
c 27	56	11.0	385	2	US-08-387-942C-24	Sequence 24, Appl
c 28	56	11.0	419	4	US-09-011-197-4	Sequence 4, Appl
c 29	56	10.5	488	2	US-08-928-692-10	Sequence 10, Appl
c 30	56	10.5	488	4	US-09-339-972-10	Sequence 10, Appl
c 31	56	11.0	496	2	US-08-949-637-2	Sequence 2, Appl
c 32	56	11.0	496	4	US-09-291-488-2	Sequence 2, Appl
c 33	56	11.0	696	1	US-08-765-081-5	Sequence 5, Appl
c 34	56	11.0	696	3	US-09-098-082-5	Sequence 5, Appl
c 35	56	11.0	696	5	PCT-US95-06994-5	Sequence 5, Appl
c 36	56	11.0	703	5	PCT-US95-06994-8	Sequence 8, Appl
c 37	56	11.0	1403	2	US-08-387-942C-3	Sequence 3, Appl
c 38	55.5	10.4	94	2	US-08-828-008-2	Sequence 2, Appl
c 39	55.5	10.4	246	4	US-09-185-160-7	Sequence 7, Appl
c 40	55.5	10.4	2186	2	US-08-822-445-2	Sequence 2, Appl
c 41	55.5	10.4	2186	4	US-09-396-540-2	Sequence 2, Appl
c 42	55	10.8	385	2	US-08-387-942C-26	Sequence 26, Appl
c 43	55	10.8	482	4	US-09-430-564-2	Sequence 2, Appl
c 44	55	10.8	495	4	US-09-430-564-3	Sequence 3, Appl
c 45	55	10.8	997	2	US-08-387-942C-4	Sequence 4, Appl

# ALIGNMENTS

RESULT 1  
US-09-251-645-13  
; Sequence 13, Application US/09251645  
; Patent No. 6281413  
; GENERAL INFORMATION:  
; APPLICANT: Krammer, Vance C.  
; APPLICANT: Morgan, Michael K.  
; APPLICANT: Anderson, Arne R.  
; APPLICANT: Hart, Hope  
; APPLICANT: Warren, Gregory W.  
; APPLICANT: Dunn, Martha  
; APPLICANT: Chen, Jeng S.  
; TITLE OF INVENTION: NOVEL INSECTICIDAL TOXINS FROM PHOTORHABDUS LUMINESCENS  
; FILE REFERENCE: CGC1963/A  
; CURRENT APPLICATION NUMBER: US/09/251,645  
; CURRENT FILING DATE: 1999-02-17  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 13  
; LENGTH: 2522  
; TYPE: PRT  
; ORGANISM: Photorhabdus luminescens  
US-09-251-645-13

Alignment Scores:  
Pred. No.: 5,38e-23 Length: 2522  
Score: 219.00 Matches: 50  
Percent Similarity: 75.2% Conservatives: 20  
Best Local Similarity: 53.76% Mismatches: 23  
Query Match: 43.20% Indels: 1  
DB: 4 Gaps: 0

US-09-856-221-3 (1-291) x US-09-251-645-13 (1-2522)

OY 1 AATACCTTGTCTCAACATTACTTACACGGCAGGATCGACAGCACTGCGACAAATTCGTCGAA 60

```
Db 2044 SerThrLeuGlnAsnIleLeuGlnArgGlnAspAlaGluAlaLeuAsnAlaLeuLeuGln 2063
QY 61 ACTCAAGCAGTGAATAGCTTTGCAGAGTATAAATGCAGGC-AAGATGATTGCTGAA 119
Db 2064 AsnGlnAlaGluLeuIleLeuThrAsnLeuSerIleGlnAspLysThrIleGluGlu 2083
QY 120 ATTGATGCTGATGAAGTGGCGCTTAAGGAACCGCTCATGCTGCACAACTCTCTTTGAC 179
Db 2084 LeuAspAlaGluLysThrValLeuGluLysSerLysAlaGlnSerArgpneasp 2103
QY 180 AGCTTCAGTACGCTGTATGACGAAGATGTTAACTCCGGTGAAACAAACGATGGATCGT 239
Db 2104 SerTySerLysLeuHisAspGluAsnIleAsnAlaGlyGluAsnGlnAlaMetThrLeu 2123
QY 240 TATCTCTCTTCATCGGTATTGACACACGACGACGAGTACGGC 278
Db 2124 ArgAlaSerAlaAlaGlyLeuThrThrAlaValGlnAla 2136
```

## RESULT 2

US-09-222-938A-34

; Sequence 34, Application US/09222938A

; Patent No. 6437108

; GENERAL INFORMATION:

; APPLICANT: Youngman, Philip

; APPLICANT: Fritz, Christian

; APPLICANT: Murphy, Christopher

; APPLICANT: Guzman, Luz-Maria

; TITLE OF INVENTION: ESSENTIAL BACTERIAL GENES AND THEIR USE

; FILE REFERENCE: 07334/060001

; CURRENT APPLICATION NUMBER: US/09/222,938A

; CURRENT FILING DATE: 1998-12-30

; NUMBER OF SEQ ID NOS: 102

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 34

; LENGTH: 299

; TYPE: PRT

; ORGANISM: Streptococcus pneumoniae

US-09-222-938A-34

```
Alignment Scores:
Pred. No.: 0.768 Length: 299
Score: 65.00 Matches: 25
Percent Similarity: 45.56% Conservative: 16
Best Local Similarity: 27.78% Mismatches: 39
Query Match: 12.82% Indels: 10
DB: 4 Gaps: 2
```

US-09-856-221-3 (1-291) x US-09-222-938A-34 (1-299)

```
QY 36 AGAAGCAGTGCAGAAATGCTGCAACTCAAGCAGTGAATAGCTTTGCAGAGTATAA 95
Db 176 GlnSerMetAsnGluIleAsnAlaGlnArgLysArgValAlaAlaGlnGluLeuAla 195
QY 96 AATGCAGGACAGATGATTGCTGGAATTTGATGCTGAAGTGGCGCTTAAGGAACCGCT 155
Db 196 GluAlaAspLysIleLysIleValThrAlaAlaGluAlaGluLysAspArgLeu 215
QY 156 CATGCT-----GCACAACTCTCT-----TTTGACAGCTTC 185
Db 216 HisGlyValGlyIleAlaGlnGlnArgLysAlaIleValAspGlyLeuAlaGluSerIle 235
QY 186 AGTACGCTGTATGACGAAGATGTTAACTCCGGTGAAACAAACGATGGATCTGTATCTC 245
Db 236 ThrGluLeuLysGluAlaAsnValGlyMetThrGluGluGlnIleMetSerIleLeuLeu 255
QY 246 TCTTCATCGGTATTGACACACGACGACGACGACGACGACGACGACGACGACGACGAC 275
Db 256 ThrAsnGlnTyrLeuAspThrLeuAsnThr 265
```

## RESULT 3

US-08-415-593-43

; Sequence 43, Application US/08415593

```
; Patent No. 5912140
; Patent No. 5912140 5776726
; GENERAL INFORMATION:
; APPLICANT: Whoriskey, Susan K.
; APPLICANT: Quinn, Cheryl L.
; APPLICANT: Tao, Nijun
; APPLICANT: Politis-Virk, Karen I.
; APPLICANT: Schimmel, Paul R.
; TITLE OF INVENTION: Recombinant Pneumocystis Aminoacyl tRNA
; TITLE OF INVENTION: Synthetase Genes, Tester Strains and Assays
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: U.S.
; ZIP: 02173
```

## COMPUTER READABLE FORM:

```
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/415,593
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Brook, David E.
; REGISTRATION NUMBER: 22,592
; REFERENCE/DOCKET NUMBER: CPI94-09
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-862-9540
; INFORMATION FOR SEQ ID NO: 43:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 978 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-415-593-43
```

```
Alignment Scores:
Pred. No.: 1.75 Length: 978
Score: 64.00 Matches: 29
Percent Similarity: 49.44% Conservative: 15
Best Local Similarity: 32.58% Mismatches: 33
Query Match: 12.62% Indels: 12
DB: 2 Gaps: 3
```

US-09-856-221-3 (1-291) x US-08-415-593-43 (1-978)

```
QY 13 AACATTACTGACGAGGATGCAGAACGACTGGCAGAA---TTGCTGCAAACTCAAGCC 69
Db 455 AsnPheMetGluLysPheProAlaAspPheIleAlaGluGlyLeuValGlnThrArgGly 474
QY 70 AGTCAATAGCTTTGCAGAGAGTATTAATGCAAGGCAAGATGATTGCTGAAATTTGATGCTG 129
Db 475 TrpPheTyrThrLeuLeu-----ValLeuGlyValGlnLeu 486
QY 130 ATCAAGTGGCGCTTAAGGAACCGCTCATGCTGCACATCTCTTTTGACAGCTTCAGTA 189
Db 487 PheGlyIleAlaProPheLysAsnValIleValAsnGlyLeuValLeuAlaSerAspGly 506
QY 190 CGCTGTATGACGAAGATGTTAACTCCGGTGAAACAAACAGGATGGATCTGTATCTCTCT 249
Db 507 LysLysMetSerLysArgLeu-----LysAsnTyrProGluLeuSerIleValLeu 523
QY 250 CATCGGTATTGACACACGACGACGACGACGACGACGACGACGACGACGACGACGACG 276
Db 524 AspLysTyrGlyAlaAspAlaLeuArg 532
```

## RESULT 4



```

; GENERAL INFORMATION:
; APPLICANT: BECKER, JEFFREY W.
; TITLE OF INVENTION: ISOLATED CANDIDA ALBICANS OLIGOPEPTIDE TRANSPORTER GENE
; FILE REFERENCE: 372.6520P
; CURRENT APPLICATION NUMBER: US/09/165,396
; CURRENT FILING DATE: 1998-10-02
; EARLIER APPLICATION NUMBER: PCT/US98/02332
; EARLIER FILING DATE: 1998-02-06
; EARLIER APPLICATION NUMBER: 60/037,859
; EARLIER FILING DATE: 1997-02-07
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 4
; LENGTH: 799
; TYPE: PRT
; ORGANISM: S. cerevisiae
; US-09-165-396-4

Alignment Scores:
Pred. No.: 3,74 Length: 799
Score: 61.50 Matches: 21
Percent Similarity: 44.44% Conservative: 15
Best Local Similarity: 25.93% Mismatches: 36
Query Match: 11.50% Indels: 9
Gaps: 4

US-09-856-221-3 (1-291) x US-09-165-396-4 (1-799)
QY 271 CTGCTGGTCTCAATFACGATGAGAGATACAGATCGCTGTTTTCACCG--- 215
Db 373 LeuTyrPheThrAsnThrTyrTyrAlaLysTyrMetProValIleSerGlySerThrTyr 392
QY 214 GAGTATACATCTCGTATACAGCGTACTGAGCTGTCAAAACGAGATTGTGCACCATGA 155
Db 393 AspAsnThrGlnAsnLysTyrAsnValThrLysIleLeuAsnGluAsp----- 408
QY 154 CGGCTTCTTAAGCGCCATCTCATCAGCATCAATTTGAGCAATCATCTTGCCTGCATT 95
Db 409 ---TyrSerIleAsnLeuGluLysTyrLysGluTyrSerProValPheValPro----- 425
QY 94 TAATCTCTGCAAGCAATTCAGTCTGAGTTTGGCAGCAATTCGCCAGTGCTCTG 35
Db 426 ---PheSerTyrLeuLeuSerTyrTyrAlaLeuAsnPheAlaIleAlaValPheVal 444
QY 34 CAT 32
Db 445 His 445

RESULT 7
5187153-4
; APPLICANT: CORDELL, BARBARA SCHILLING, JAMES W.; KATUNUMA, NOBUHIKO
; TITLE OF INVENTION: METHODS OF TREATMENT USING ALZHEIMER'S
; AMYLOID POLYPEPTIDE DERIVATIVES
; NUMBER OF SEQUENCES: 33
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/502,273
; FILING DATE: 29-MAR-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 361,912
; FILING DATE: 06-JUN-1989
; APPLICATION NUMBER: 359,911
; FILING DATE: 12-MAY-1989
; APPLICATION NUMBER: 87,002
; FILING DATE: 18-AUG-1987
; APPLICATION NUMBER: 8,810
; FILING DATE: 30-JAN-1987
; APPLICATION NUMBER: 948,376
; FILING DATE: 31-DEC-1986
; APPLICATION NUMBER: 932,193
; FILING DATE: 17-NOV-1986
; SEQ ID NO:4:

```

```

; LENGTH: 152
; 5187153-4

Alignment Scores:
Pred. No.: 2,65 Length: 152
Score: 60.50 Matches: 21
Percent Similarity: 37.14% Conservative: 11
Best Local Similarity: 57.50% Mismatches: 13
Query Match: 11.31% Indels: 11
Gaps: 5

US-09-856-221-3 (1-291) x 5187153-4 (1-152)
QY 233 CATCGCTGTTTTC---ACCGAGTTTAATCTTCGTATACAGCGTACTGAAGCTGTC 177
Db 12 TyrArgValPhePheLysSerHisLeuSerIlePhePheLeuGlnAlaLysGlu---Val 30
QY 176 AAAACGAGATTGTGC---ACCATGACGGCTTTCCTTAAAGCCCACTTCATCAGCATCAAT 120
Db 31 LeuGlyArgLeuCysLeuThrValLeuLeuPheMetSerTys----- 44
QY 119 TTCAGCAATCATCTGCTGCATTTTAACTACTCTGCAAGCTAATTCATCA 72
Db 45 ---ThrAsnTrpLeuSerCysIleLeu---LeuCysAsnThrGlySer 58

; RESULT 8
; US-09-034-916-13
; Sequence 13, Application US/09034916
; Patent No. 6046314
; GENERAL INFORMATION:
; APPLICANT: GEBE, JOHN A.
; APPLICANT: SIADAK, ANTHONY W.
; APPLICANT: ARUFFO, ALEJANDRO A.
; TITLE OF INVENTION: SPALPHA: A NOVEL SCAVENGER RECEPTOR
; TITLE OF INVENTION: CYSTEINE-RICH DOMAIN-CONTAINING POLYPEPTIDE AND MONOCLONAL
; TITLE OF INVENTION: ANTIBODIES THERETO
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ROBINS & ASSOCIATES
; STREET: 90 MIDDLEFIELD ROAD, SUITE 200
; CITY: MENLO PARK
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/034,916
; FILING DATE: 04-MAR-1998
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/039,956
; FILING DATE: 06-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: ROBINS, ROBERTA L.
; REGISTRATION NUMBER: 33,208
; REFERENCE/DOCKET NUMBER: 5998-0020
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 325-7812
; TELEFAX: (650) 325-7823
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 100 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-09-034-916-13

Alignment Scores:

```



```
Percent Similarity: 41.12% Conservative: 15
Best Local Similarity: 27.10% Mismatches: 38
Query Match: 10.93% Indels: 25
DB: 4 Gaps: 5

US-09-856-221-3 (1-291) x US-09-024-472-2 (1-169)

QY 286 ATATGCGGCGCTACTGCTGCTCAATACCGATGAGAGAGATACAGA----- 236
Db 24 IleaGAsnAlaValProLeuGlyThrThraLaLysGluGluMetGluArgPheTrpAsn 43
QY 235 -----TCCATCGCTTGTGTTTTCACCGAGTTAAACATCTTCGTACATACACGGA 188
Db 44 LysAsnIleGlySerAsnArgProLeuSerProHisIleThrIleTyrSerTrpSerLeu 63
QY 187 -----CTCAAGCTCTCAAAACGAGATTGTGCACCATGACGGCTTTCCTTAAGCGCC 137
Db 64 ProMetAlaMetSerIleCysHisArgGlyThrGly-----IleAlaLeuSerAla 80
QY 136 ---ACTCATCAGCATCAATTTCAGCAATCATCTGCGCT----- 101
Db 81 GlyValSerLeuPheGlyMetSerAlaLeuLeuLeuProGlyAsnPheGluSerTyrLeu 100
QY 100 -----GCATTTAATACTCTGCAAAAGCTAATTCACTGCTTGTGAGTTTGCAGCA 53
Db 101 GluLeuValLysSerLeuCysLeuGlyProAlaLeuIleHisThrAlaLysPheAlaLeu 120
QY 52 ATTCTGCCAGTCTTCTGCAT 32
Db 121 ValPheProLeuMetTyrHis 127

RESULT 11
US-09-543-084A-32
; Sequence 32, Application US/09543084A
; Patent No. 6361988
; GENERAL INFORMATION:
; APPLICANT: Frances H. Arnold
; APPLICANT: Zhixin Shao
; APPLICANT: Huimin Zhao
; APPLICANT: Lorraine J. Giver
; TITLE OF INVENTION: Recombination of Polynucleotide
; TITLE OF INVENTION: Sequences Using Defined or Random Primer Sequences
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Oppenheimer Wolff & Donnelly LLP
; STREET: 2029 Century Park East, Suite 3800
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90067
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Windows 98
; SOFTWARE: Word 97
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/543,084A
; FILING DATE: April 4, 2000
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/381,935
; FILING DATE: March 25, 1997
; APPLICATION NUMBER: 60/041,666
; FILING DATE: April 30, 1997
; APPLICATION NUMBER: 60/045,211
; FILING DATE: May 12, 1997
; APPLICATION NUMBER: 60/046,256
; FILING DATE: August 4, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Oldenkamp, David J.
; REGISTRATION NUMBER: 29,421
; REFERENCE/DOCKET NUMBER: 330187-89

TELECOMMUNICATION INFORMATION:
; TELEPHONE: (310) 788-5000
; TELEFAX: (310) 788-5100
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 782 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-543-084A-32
Alignment Scores:
Pred. No.: 10.2 Length: 782
Score: 58.50 Matches: 17
Percent Similarity: 50.00% Conservative: 5
Best Local Similarity: 38.64% Mismatches: 17
Query Match: 10.93% Indels: 5
DB: 4 Gaps: 1

US-09-856-221-3 (1-291) x US-09-543-084A-32 (1-782)
QY 229 GCTTGTTTTCACCGGAGTTAAACATCTTGTCTATACACGCTACTGAAGCTGTCAAACGA 170
Db 440 AlaProPheGlnProLeuTyrAlaSerSerGlyGlnAlaValLeuAspGlySerArgSer 459
QY 169 GATTGTGCGACCATGACGGCTTTCCTTTAAGCGGCACCTTCATCAGCATCAATTTTCAGCAATC 110
Db 460 AspCysAla-----LeuGlyAlaAspProAspAlaIleValProGlyIle 474
QY 109 ATCTTGCCTGCA 98
Db 475 LeuGlyProAla 478

RESULT 12
US-09-543-084A-33
; Sequence 33, Application US/09543084A
; Patent No. 6361988
; GENERAL INFORMATION:
; APPLICANT: Frances H. Arnold
; APPLICANT: Zhixin Shao
; APPLICANT: Huimin Zhao
; APPLICANT: Lorraine J. Giver
; TITLE OF INVENTION: Recombination of Polynucleotide
; TITLE OF INVENTION: Sequences Using Defined or Random Primer Sequences
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Oppenheimer Wolff & Donnelly LLP
; STREET: 2029 Century Park East, Suite 3800
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90067
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Windows 98
; SOFTWARE: Word 97
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/543,084A
; FILING DATE: April 4, 2000
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/381,935
; FILING DATE: March 25, 1997
; APPLICATION NUMBER: 60/041,666
; FILING DATE: April 30, 1997
; APPLICATION NUMBER: 60/045,211
; FILING DATE: May 12, 1997
; APPLICATION NUMBER: 60/046,256
; FILING DATE: August 4, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Oldenkamp, David J.
; REGISTRATION NUMBER: 29,421
; REFERENCE/DOCKET NUMBER: 330187-89
```







GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: January 15, 2003, 15:46:43 ; Search time 5.45922 Seconds  
(without alignments)  
2119.222 Million cell updates/sec

Title: US-09-856-221-3

Perfect score: 507

Sequence: 1 aataactgtcacaattac.....taagccctgcatagtgctc 291

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 120991 seqs, 19878514 residues

Total number of hits satisfying chosen parameters: 241982

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DRV=xlp  
-O=/cgn2\_1/USFT0\_spool/US09856221/runat\_15012003\_153926\_12555/app\_query.fasta\_1.1948  
-DB=PublishedApplications\_AA -QPM=fastan -SUFFIX=rapb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR\_SCORE=pct -THR\_MAX=100  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=200000000 -USER=US09856221@cgn\_1\_1\_14 @runat\_15012003\_153926\_12555  
-NCPU=6 -ICPU=3 -NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FCGAPOP=6  
-FCGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications\_AA.\*  
1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pcp.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pcp.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pcp.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pcp.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pcp.\*  
6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pcp.\*  
7: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pcp.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pcp.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pcp.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pcp.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pcp.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pcp.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pcp.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pcp.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Length	ID	Description
1	220	43.4	2516	10 US-09-817-514A-2
2	208	41.0	2504	10 US-09-817-514A-8
3	65.5	12.9	706	9 US-09-738-626-6019
4	65	12.8	508	10 US-09-833-745-40

5	62.5	12.3	513	10 US-09-833-745-49	Sequence 49, Appl
6	62	12.2	126	10 US-09-925-299-1486	Sequence 1486, Ap
c 7	61	11.4	296	10 US-09-771-730-116	Sequence 116, App
c 8	61	11.4	296	10 US-09-771-730-120	Sequence 120, App
c 9	61	11.4	303	10 US-09-771-730-118	Sequence 118, App
c 10	61	11.4	332	10 US-09-771-730-28	Sequence 28, Appl
c 11	61	11.4	994	10 US-09-852-909-2	Sequence 2, Appl
12	61	12.0	1319	9 US-10-042-431-14	Sequence 14, Appl
13	61	12.0	1413	9 US-10-042-431-13	Sequence 13, Appl
14	61	12.0	1453	9 US-10-042-431-11	Sequence 11, Appl
15	61	12.0	2167	10 US-09-801-368-56	Sequence 56, Appl
c 16	59.5	11.1	162	10 US-09-747-155-227	Sequence 227, App
17	59	11.6	144	10 US-09-925-301-1538	Sequence 1538, Ap
c 18	59	11.0	313	10 US-09-886-055-461	Sequence 461, App
19	59	11.6	1149	9 US-09-977-577-10	Sequence 10, Appl
20	59	11.6	1149	9 US-09-977-577-11	Sequence 11, Appl
21	59	11.6	1151	9 US-09-977-577-13	Sequence 13, Appl
22	59	11.6	1156	9 US-09-977-577-12	Sequence 12, Appl
c 23	58.5	10.9	787	10 US-09-205-448-8	Sequence 8, Appl
c 24	58	11.4	350	10 US-09-881-752A-234	Sequence 234, App
25	58	11.4	464	10 US-09-863-824-2	Sequence 2, Appl
c 26	58	10.8	550	9 US-09-738-626-5417	Sequence 5417, Ap
c 27	57.5	10.7	756	10 US-09-919-497-66	Sequence 66, Appl
c 28	57	10.7	56	10 US-09-764-869-1049	Sequence 1049, Ap
c 29	57	11.2	250	10 US-09-815-242-10871	Sequence 10871, A
c 30	56.5	10.6	94	10 US-09-811-284-203	Sequence 203, App
c 31	56.5	10.6	209	10 US-09-393-634-13	Sequence 13, Appl
32	56.5	11.1	241	10 US-09-764-864-936	Sequence 936, App
c 33	56.5	10.6	324	10 US-09-886-055-217	Sequence 217, App
c 34	56.5	10.6	619	9 US-09-895-913A-326	Sequence 326, App
35	56.5	11.1	1164	10 US-09-950-046A-2	Sequence 2, Appl
c 36	55.5	10.9	94	10 US-09-764-864-1558	Sequence 1558, Ap
c 37	55.5	10.4	94	10 US-09-739-707-2	Sequence 2, Appl
c 38	55.5	10.4	928	9 US-10-108-605-261	Sequence 261, App
c 39	55.5	10.9	969	10 US-09-118-276-2	Sequence 2, Appl
40	55.5	10.9	1354	10 US-09-808-571A-4	Sequence 4, Appl
41	55.5	10.4	1447	10 US-09-808-571A-2	Sequence 2, Appl
c 42	55.5	10.9	2186	10 US-09-927-668-2	Sequence 2, Appl
43	55	10.8	378	10 US-09-864-761-43251	Sequence 43251, A
44	55	10.8	469	10 US-09-764-864-1136	Sequence 1136, Ap
c 45	55	10.3	876	9 US-09-738-626-4789	Sequence 4789, Ap

ALIGNMENTS

RESULT 1  
US-09-817-514A-2  
; Sequence 2, Application US/09817514A  
; Patent No. US20020078478A1  
; GENERAL INFORMATION:  
; APPLICANT: ffrench-Constant, Richard  
; APPLICANT: Bowen, David  
; APPLICANT: Rocheleau, Thomas  
; APPLICANT: Waterfield, Nicholas  
; TITLE OF INVENTION: DNA SEQUENCES FROM PHOTORHABDUS LUMINESCENS  
; FILE REFERENCE: 61645  
; CURRENT APPLICATION NUMBER: US/09/817,514A  
; CURRENT FILING DATE: 2000-03-26  
; PRIOR APPLICATION NUMBER: US 60/191806  
; PRIOR FILING DATE: 2000-03-24  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 2  
; LENGTH: 2516  
; TYPE: PRT  
; ORGANISM: Photorhabdus luminescens  
US-09-817-514A-2

Alignment Scores:  
Pred. No.: 2,14e-21 Length: 2516  
Score: 220.00 Matches: 50  
Percent Similarity: 74.1% Conservative: 19  
Best Local Similarity: 53.7% Mismatches: 24

```
Query Match: 43.39% Indels: 1
DB: 10 Gaps: 0
US-09-856-221-3 (1-291) x US-09-817-514A-2 (1-2516)
QY 1 AATACCTTGCCTCAACATTACTGAACGCGAGGATGCGAGCAAGCACTGGCAGAAATTCGTGCAA 60
Db 2038 SerThrLeuGlnAsnIleIleGluArgGlnAspAlaGluAlaLeuAsnAlaLeuGln 2057
QY 61 ACTCAAGCAGCTGAATAGCTTTCAGAGCTATTAAATGACAGGC-AAAGATGATTGCTGAA 119
Db 2058 AsnGlnAlaGluLeuIleLeuThrAsnLeuSerIleGlnAspLysThrIleGluGlu 2077
QY 120 ATGTAGCTGTAGTAAGTGGCGCTTAAGGAAAGCCGTCATGTTGCGACAAATCTCGTTTGGAC 179
Db 2078 LeuAspAlaGluLysThrValLeuGluLysSerLysAlaGluSerArgPheAsp 2097
QY 180 AGCTTCAGTACGCTGTATGACGAAGATGTTAACTCCGGTGAAGAAACAAAGCGATGATCG 239
Db 2098 SerTyrGlyLysLeuTyrAspGluAsnIleAsnAlaGlyGluAsnGlnAlaMetThrLeu 2117
QY 240 TATCTCTCTTCATCGGTATTGACGACGACGATGACGGC 278
Db 2118 ArgAlaSerAlaAlaGlyLeuThrAlaValGlnAla 2130
RESULT 2
US-09-817-514A-8
; Sequence 8, Application US/09817514A
; Patent No. US20020078478A1
; GENERAL INFORMATION:
; APPLICANT: firench-Constant, Richard
; APPLICANT: Bowen, David
; APPLICANT: Rocheleau, Thomas
; APPLICANT: Waterfield, Nicholas
; TITLE OF INVENTION: DNA SEQUENCES FROM PHOTORHABDUS LUMINESCENS
; FILE REFERENCE: 61645
; CURRENT APPLICATION NUMBER: US/09/817,514A
; PRIOR FILING DATE: 2000-03-26
; PRIOR APPLICATION NUMBER: US 60/191806
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
; LENGTH: 2504
; TYPE: PRT
; ORGANISM: Photorhabdus luminescens
US-09-817-514A-8
Alignment Scores:
Pred. No.: 9,16e-20 Length: 2504
Score: 208.00 Matches: 46
Percent Similarity: 72.92% Conservative: 24
Best Local Similarity: 47.92% Mismatches: 26
Query Match: 41.03% Indels: 1
DB: 10 Gaps: 0
US-09-856-221-3 (1-291) x US-09-817-514A-8 (1-2504)
QY 1 AATACCTTGCCTCAACATTACTGAACGCGAGGATGCGAGCAAGCACTGGCAGAAATTCGTGCAA 60
Db 2032 SerSerLeuLeuGlyTyrSerGluArgGlnAspAlaGluAlaMetSerGlnLeuLeuGln 2051
QY 61 ACTCAAGCAGCTGAATAGCTTTCAGAGCTATTAAATGACAGGC-AAAGATGATTGCTGAA 119
Db 2052 ThrGlnAlaSerGluLeuIleLeuThrSerIleArgMetGlnAspAsnGlnLeuAlaGlu 2071
QY 120 ATGTAGCTGTAGTAAGTGGCGCTTAAGGAAAGCCGTCATGTTGCGACAAATCTCGTTTGGAC 179
Db 2072 LeuAspSerGluLysThrAlaLeuGlnValSerLeuAlaGlyValGlnGlnArgPheAsp 2091
QY 180 AGCTTCAGTACGCTGTATGACGAAGATGTTAACTCCGGTGAAGAAACAAAGCGATGATCG 239
Db 2092 SerTyrSerGlnLeuTyrGluAsnIleAsnAlaGlyGluGlnArgAlaLeuAlaLeu 2111
QY 240 TATCTCTCTTCATCGGTATTGACGACGACGATGACGGC 287
Db 2112 ArgSerGluSerAlaIleGluSerGlnGlyAlaGlnIleSerArgMet 2127
RESULT 3
US-09-738-626-6019
; Sequence 6019, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 6019
; LENGTH: 706
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-6019
Alignment Scores:
Pred. No.: 1.46 Length: 706
Score: 65.50 Matches: 15
Percent Similarity: 48.28% Conservative: 13
Best Local Similarity: 25.86% Mismatches: 23
Query Match: 12.92% Indels: 7
DB: 9 Gaps: 1
US-09-856-221-3 (1-291) x US-09-738-626-6019 (1-706)
QY 132 GAATGGCGCTTAAGGAAAGCCGTCATGCTGCACAAATCTCGTTTTCACAGCTTCAGTACG 191
Db 335 GluThrAlaGlnSerGluSerValHisGlyThrGluProAspArgAspGluLeuThrMet 354
QY 192 CTGTAT-----GACGAAGATGTTAACTCCGGTGAAGAAACAAAGCG 230
Db 355 PheTyrMetTrpLeuGlnTrpLeucysaspGluGlnIleuAlaAlaGlnIlysArgAla 374
QY 231 ATGGATCTGATCTCTTCATCGGTATTGACGACGACGATGACGGCCCTGCAT 284
Db 375 ValAspAlaGlyMetSerIleGlyIleMetAlaAspLeuAlaValGlyValHis 392
RESULT 4
US-09-833-745-40
; Sequence 40, Application US/09833745
; Patent No. US20020052038A1
; GENERAL INFORMATION:
; APPLICANT: ROBERTS, JOSEPH
; APPLICANT: SETHURAMAN, NATARAJAN
; APPLICANT: MACALISTER, THOMAS
; TITLE OF INVENTION: CLONING, OVEREXPRESSION AND THERAPEUTIC USE OF
; FILE REFERENCE: 078728/0106
; CURRENT APPLICATION NUMBER: US/09/833,745
; CURRENT FILING DATE: 2001-04-13
```

```

; PRIOR APPLICATION NUMBER: 60/197,770
; PRIOR FILING DATE: 2000-04-14
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 40
; LENGTH: 508
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US-09-833-745-40

```

Alignment Scores:	1.54	Length:	508
Pred. No.:	Score:	Matches:	22
	65.00	Conservative:	16
	Percent Similarity:	Mismatches:	18
	51.3%	Indels:	4
	Best Local Similarity:	Gaps:	18
	29.7%		
	Query Match:		
	12.83%		
	DB:		
	10		

US-09-856-221-3 (1-291) x US-09-833-745-40 (1-508)

QY	108	ATGATTGCTGAATATGATGCTGATGATGAAGTGGCGCTTAAAGGAACCGCT	-----	155
DB	204	valvalalataltyrilleglualagluLysLeuallatyrGlnThrGluarGilleAlaSerLeu		223
QY	156	-----CATGTGTCACAACTCTCGCTTTTGACACGCTTCACTAGTCGCTGTATGACGAAGATGTT	209	
DB	224	ThrIleGluGlyLeuGlnGlyIlelleAspAlaPhe		239
QY	210	AAC-----TCCGTGAAAAACAGCGATGGATCG-----TATCTC	245	
DB	240	HisLeuAlaArgGlyTyrGlnGluGlnIleaspValAlaGluarGilleArGpheTyrLeu	259	
QY	246	TCTTCATCGGTATTGACACACAGCATAGCGCCCTGCATATG	287	
DB	260	SerAspSerGlyLeuThrThrSerGlnGlyGluLeuArGVal	273	

## RESULT 5

```

US-09-833-745-49
; Sequence 49, Application US/09833745
; Patent No. US20020052038A1
; GENERAL INFORMATION:
; APPLICANT: ROBERTS, JOSEPH
; APPLICANT: SETHURAMAN, NATARAJAN
; APPLICANT: MACALLISTER, THOMAS
; TITLE OF INVENTION: CLONING, OVEREXPRESSION AND THERAPEUTIC USE OF
; TITLE OF INVENTION: BIOACTIVE HISTIDINE AMMONIA LYASE
; FILE REFERENCE: 078728/0106
; CURRENT APPLICATION NUMBER: US/09/833,745
; PRIOR FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: 60/197,770
; PRIOR FILING DATE: 2000-04-14
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 49
; LENGTH: 513
; TYPE: PRT
; ORGANISM: Bacillus subtilis
; US-09-833-745-49

```

Alignment Scores:		
Pred. No.:	3.38	Length: 513
Score:	62.50	Matches: 21
Percent Similarity:	52.17%	Conservative: 15
Best Local Similarity:	30.43%	Mismatches: 14
Query Match:	12.33%	Indels: 19
DB:	10	Gaps: 4

US-09-856-221-3 (1-291) x US-09-833-745-49 (1-513)

QY	155
108 ATGATTGCTGAATTCATGCTGATGAAGTGGCGCTTAAGGAACCCGT-----	
:	::::     :::     :::
D6	229
210 ValValAlaTyrlleGluAlaGluLysLeuAlaTyrglnThrGluArgIleAalaSerLeu	

Qy	156	-----CATGTGCACAATCTCGTTTTCAGACGCTTCAGTCCCTGTATGACGAAGATGTT
Dd	230	ThrlleGlucgluYleugluGlylleIleAspAlaphe-----ASPLUASPille
Qy	210	AAC-----TCCGGTGAAAAAACAAGCGATCGATCTG-----TAT
Dd	246	HlLeuAlalaLeuargGlyTyrrgluGluInileAspValalagluargileargPheTyr
Qy	243	CTCTCTTCATCGGTATTGACGACCAGC
Dd	266	LeuSerAspSerGlyLeuThrThrSer

## RESULT 6

```

US-09-925-299-1486
; Sequence 1486, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PAL02
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05983
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1486
; LENGTH: 126
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (112)
; OTHER INFORMATION: xaa equals any of the naturally occurring
US-09-925-299-1486

```

Alignment Scores:		
pred. No.:	2.53	Length:
Score:	62.00	Matches:
Percent Similarity:	73.33%	Conservative:
Best Local Similarity:	66.67%	Mismatches:
Query Match:	12.23%	Indels:
DB:	10	Gaps:
		0
		0
		10
		126

US-09-856-221-3 (1-291) x US-09-925-299-1486 (1-126)

5 CCTTGCTCAACATTACTGAACGGCAGGATGCAGAAGCACTGGCAG 49  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 34 ProCysSerLeuLeuLeuArgGlyArgValThrSerHisTrpGln 48

## RESULT 7

```

US-09-771-730-116
: Sequence 116, Application US/09771730
: Patent No. US20020146807A1
: GENERAL INFORMATION:
: APPLICANT: Prayaga, Sudhirdas K.
: APPLICANT: Li, Li
: APPLICANT: Padigaru, Muralidhara
: APPLICANT: MacDougall, John R.
: APPLICANT: Spytek, Kimberly Ann
: APPLICANT: Tchernev, Velizar T.
: APPLICANT: Vernet, Corine A. M.
: TITLE OF INVENTION: NOVEL POLYPEPTIDES AND NUCLEIC ACIDS ENCODING SAME
: FILE REFERENCE: 15966-645
: CURRENT APPLICATION NUMBER: US/09/771,730
: PRIOR FILING DATE: 2001-08-21
: CURRENT APPLICATION NUMBER: 60/178,413
: PRIOR FILING DATE: 2000-01-27
: PRIOR APPLICATION NUMBER: 60/178,371
: PRIOR FILING DATE: 2000-01-27
: PRIOR APPLICATION NUMBER: 60/178,408

```

; PRIOR FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: 60/178,370  
; PRIOR FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: 60/178,406  
; PRIOR FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: 60/178,414  
; PRIOR FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: 60/178,409  
; PRIOR FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: 60/178,634  
; PRIOR FILING DATE: 2000-02-07  
; PRIOR APPLICATION NUMBER: 60/220,516  
; PRIOR FILING DATE: 2000-07-24  
; PRIOR APPLICATION NUMBER: 60/221,408  
; PRIOR FILING DATE: 2000-07-28  
; PRIOR APPLICATION NUMBER: 60/221,943  
; PRIOR FILING DATE: 2000-07-31  
; PRIOR APPLICATION NUMBER: 60/257,599  
; PRIOR FILING DATE: 2000-12-21  
; PRIOR APPLICATION NUMBER: 60/260,290  
; NUMBER OF SEQ ID NOS: 148  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 116  
; LENGTH: 296  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-771-730-116

Alignment Scores:  
Pred. No.: 4.53 Length: 296  
Score: 61.00 Matches: 22  
Percent Similarity: 42.86% Conservative: 14  
Best Local Similarity: 26.19% Mismatches: 26  
Query Match: 11.40% Indels: 23  
DB: Gaps: 3

US-09-856-221-3 (1-291) x US-09-771-730-116 (1-296)

QY 288 CCATATGCGAGCGCG-----TACTGCTGGTCTCAATACCG---ATGAAGAGA 244  
|||::: ||||| ||| :::: ||||| ||| |||  
Db 161 ProPheCysGlyProAsnArgValAspTyrIlePheCysAspIleProAlaMetLeuArg 180  
QY 243 GA-----TACAGA 236  
Db 181 LeuAlaCysAlaAspThrAlaIleAsnGluLeuValThrPheAlaAspIleGlyPheLeu 200  
QY 235 TCATCGCTTGTTTTCACCGGAGTTAACAATCTTCGTCATACAGCGTACTGAAGCTGTCA 176  
::: :::: ||||| ||| ||||| ||| |||  
Db 201 AlaLeuThrCysPheMetProIleLeuThrSerTyrGlyTyrIleValAlaAlaIleLeu 220  
QY 175 AAACGAGATTGTGCACCATGACGCTTTCCTTAAGCGGCACCTTCATCAGCATCAATTCA 116  
::: ||| :::: ||| :::: ||| :::: ||| :::: |||  
Db 221 ArgIleProSerAlaAspGlyArgAsnAlaPheSerThrCysAlaAlaHisLeuThr 240  
QY 115 GCAATCATCTTG 104  
::: ||| :::: ||| :::: ||| :::: |||  
Db 241 ValValIleVal 244

RESULT 8  
US-09-771-730-120  
; Sequence 120, Application US/09771730  
; Patent No. US20020146807A1  
; GENERAL INFORMATION:  
; APPLICANT: Prayaga, Sudhirdas K.  
; APPLICANT: Li, Li  
; APPLICANT: Padigar, Muralidhara  
; APPLICANT: MacDougall, John R.  
; APPLICANT: Spytek, Kimberly Ann  
; APPLICANT: Tchernev, Velizar T.  
; APPLICANT: Vernet, Corine A. M.  
; TITLE OF INVENTION: NOVEL POLYPEPTIDES AND NUCLEIC ACIDS ENCODING SAME  
; FILE REFERENCE: 15966-645

; CURRENT APPLICATION NUMBER: US/09/771,730  
; CURRENT FILING DATE: 2001-08-21  
; PRIOR APPLICATION NUMBER: 60/178,413  
; PRIOR FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: 60/178,371  
; PRIOR FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: 60/178,408  
; PRIOR FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: 60/178,370  
; PRIOR FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: 60/178,406  
; PRIOR FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: 60/178,414  
; PRIOR FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: 60/178,409  
; PRIOR FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: 60/180,634  
; PRIOR FILING DATE: 2000-02-07  
; PRIOR APPLICATION NUMBER: 60/220,516  
; PRIOR FILING DATE: 2000-07-24  
; PRIOR APPLICATION NUMBER: 60/221,408  
; PRIOR FILING DATE: 2000-07-28  
; PRIOR APPLICATION NUMBER: 60/221,943  
; PRIOR FILING DATE: 2000-07-31  
; PRIOR APPLICATION NUMBER: 60/257,599  
; PRIOR FILING DATE: 2000-12-21  
; PRIOR APPLICATION NUMBER: 60/260,290  
; NUMBER OF SEQ ID NOS: 148  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 120  
; LENGTH: 296  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-771-730-120

Alignment Scores:  
Pred. No.: 4.53 Length: 296  
Score: 61.00 Matches: 22  
Percent Similarity: 42.86% Conservative: 14  
Best Local Similarity: 26.19% Mismatches: 26  
Query Match: 11.40% Indels: 23  
DB: Gaps: 3

US-09-856-221-3 (1-291) x US-09-771-730-120 (1-296)

QY 288 CCATATGCGAGCGCG-----TACTGCTGGTCTCAATACCG---ATGAAGAGA 244  
|||::: ||||| ||| :::: ||||| ||| |||  
Db 161 ProPheCysGlyProAsnArgValAspTyrIlePheCysAspIleProAlaMetLeuArg 180  
QY 243 GA-----TACAGA 236  
Db 181 LeuAlaCysAlaAspThrAlaIleAsnGluLeuValThrPheAlaAspIleGlyPheLeu 200  
QY 235 TCATCGCTTGTTTTCACCGGAGTTAACAATCTTCGTCATACAGCGTACTGAAGCTGTCA 176  
::: :::: ||||| ||| ||||| ||| |||  
Db 201 AlaLeuThrCysPheMetProIleLeuThrSerTyrGlyTyrIleValAlaAlaIleLeu 220  
QY 175 AAACGAGATTGTGCACCATGACGCTTTCCTTAAGCGGCACCTTCATCAGCATCAATTCA 116  
::: ||| :::: ||| :::: ||| :::: ||| :::: |||  
Db 221 ArgIleProSerAlaAspGlyArgAsnAlaPheSerThrCysAlaAlaHisLeuThr 240  
QY 115 GCAATCATCTTG 104  
::: ||| :::: ||| :::: ||| :::: |||  
Db 241 ValValIleVal 244

RESULT 9  
US-09-771-730-118  
; Sequence 118, Application US/09771730  
; Patent No. US20020146807A1  
; GENERAL INFORMATION:  
; APPLICANT: Prayaga, Sudhirdas K.  
; APPLICANT: Li, Li

```

; APPLICANT: Padigaru, Muralidhara
; APPLICANT: MacDougall, John R.
; APPLICANT: Spytek, Kimberly Ann
; APPLICANT: Tchernev, Velizar T.
; APPLICANT: Vernet, Corine A. M.
; TITLE OF INVENTION: NOVEL POLYPEPTIDES AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 15966-645
; CURRENT APPLICATION NUMBER: US/09/771,730
; PRIOR FILING DATE: 2001-08-21
; PRIOR APPLICATION NUMBER: 60/178,413
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: 60/178,371
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: 60/178,408
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: 60/178,370
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: 60/178,406
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: 60/178,414
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: 60/178,409
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: 60/180,634
; PRIOR FILING DATE: 2000-02-07
; PRIOR APPLICATION NUMBER: 60/220,516
; PRIOR FILING DATE: 2000-07-24
; PRIOR APPLICATION NUMBER: 60/221,408
; PRIOR FILING DATE: 2000-07-28
; PRIOR APPLICATION NUMBER: 60/221,943
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: 60/257,599
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: 60/260,290
; PRIOR FILING DATE: 2001-01-08
; NUMBER OF SEQ ID NOS: 148
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 118
; LENGTH: 303
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-771-730-118

Alignment Scores:
Pred. No.: 4.57 Length: 303
Score: 61.00 Matches: 22
Percent Similarity: 42.86% Conservative: 14
Best Local Similarity: 26.19% Mismatches: 26
Query Match: 11.40% Indels: 23
DB: 10 Gaps: 3

US-09-856-221-3 (1-291) x US-09-771-730-118 (1-303)
QY 288 CCATATGCGAGCGCG-----TACTGCTGGTGCTCAATACCG---ATGAAGAGA 244
Db 167 PropheCysGlyProAsnArgValAspTyrIlePheCysAspIleProAlaMetLeuArg 186
QY 243 GA-----TACAGA 236
Db 187 LeuAlaCysAlaAspThrAlaIleAsnGluLeuValThrPheAlaAspIleGlyPheLeu 206
QY 235 TCCATCGCTGCTTTTCCACCGAGTTAAACATCTTCGTATACACGCTACTGAAGCTGTCA 176
Db 207 AlaLeuThrCysPheMetProIleLeuThrSeryrGlyrIleValAlaIleLeu 226
QY 175 AAACGAGATTGTGCACCATGCGGCTTTCCTTAAGCGCCACTTCATCAGCATCAATTCA 116
Db 227 ArgIleProSerAlaAspGlyArgAsnAlaPheSerThrCysAlaAlaHisLeuThr 246
QY 115 GCAATCATCTTG 104
Db 247 ValValIleVal 250
```

```

RESULT 10
US-09-771-730-28
; Sequence 28, Application US/09771730
; Patent No. US20020146807A1
; GENERAL INFORMATION:
; APPLICANT: Prayaga, Sudhirdas K.
; APPLICANT: Li, Li
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: MacDougall, John R.
; APPLICANT: Spytek, Kimberly Ann
; APPLICANT: Tchernev, Velizar T.
; APPLICANT: Vernet, Corine A. M.
; TITLE OF INVENTION: NOVEL POLYPEPTIDES AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 15966-645
; CURRENT APPLICATION NUMBER: US/09/771,730
; CURRENT FILING DATE: 2001-08-21
; PRIOR APPLICATION NUMBER: 60/178,413
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: 60/178,371
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: 60/178,408
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: 60/178,370
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: 60/178,406
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: 60/178,414
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: 60/178,409
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: 60/180,634
; PRIOR FILING DATE: 2000-02-07
; PRIOR APPLICATION NUMBER: 60/220,516
; PRIOR FILING DATE: 2000-07-24
; PRIOR APPLICATION NUMBER: 60/221,408
; PRIOR FILING DATE: 2000-07-28
; PRIOR APPLICATION NUMBER: 60/221,943
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: 60/257,599
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: 60/260,290
; PRIOR FILING DATE: 2001-01-08
; NUMBER OF SEQ ID NOS: 148
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 28
; LENGTH: 332
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-771-730-28

Alignment Scores:
Pred. No.: 4.7 Length: 332
Score: 61.00 Matches: 22
Percent Similarity: 42.86% Conservative: 14
Best Local Similarity: 26.19% Mismatches: 26
Query Match: 11.40% Indels: 23
DB: 10 Gaps: 3

US-09-856-221-3 (1-291) x US-09-771-730-28 (1-332)
QY 288 CCATATGCGAGCGCG-----TACTGCTGGTGCTCAATACCG---ATGAAGAGA 244
Db 188 PropheCysGlyProAsnArgValAspTyrIlePheCysAspIleProAlaMetLeuArg 207
QY 243 GA-----TACAGA 236
Db 208 LeuAlaCysAlaAspThrAlaIleAsnGluLeuValThrPheAlaAspIleGlyPheLeu 227
QY 235 TCCATCGCTGCTTTTCCACCGAGTTAAACATCTTCGTATACACGCTACTGAAGCTGTCA 176
Db 228 AlaLeuThrCysPheMetProIleLeuThrSeryrGlyrIleValAlaIleLeu 247
QY 175 AAACGAGATTGTGCACCATGAGCGGCTTTCCTTAAGCGCCACTTCATCAGCATCAATTCA 116
```









GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: January 15, 2003, 15:44:53 ; Search time 13.4107 Seconds  
(without alignments)  
4172.064 Million cell updates/sec

Title: US-09-856-221-3  
Perfect score: 507  
Sequence: 1 aatacctgtcgaacattac.....tacggccctgcataatggtgc 291

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 566448

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp  
-O=/cpn2.1/USP70.spool/US0856221/runat\_15012003\_153925\_12500/app.query.fasta\_1.1948  
-DB=PIR\_73 -QFMT=FASTAN -SUFFIX=rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=BLOSUM62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORMEXT=HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US0856221 -CGN\_1\_1\_95\_runat\_15012003\_153925\_12500 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR\_73.\*

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query	Length	ID	Description
1	69	13.6	1379	2	S37310
2	68	13.4	299	2	T32094
3	68	13.4	329	2	A69791
4	68	13.4	1383	2	T07126
5	68	13.4	1693	2	AC3240
6	67	12.5	1345	2	H71608
7	67	13.2	1380	2	S64721
8	66.5	12.4	520	2	S45753
9	66.5	12.4	585	2	T09782
10	66	13.0	218	2	A81764
11	65.5	12.2	930	2	T00403
12	65	12.1	252	2	T37633
13	65	12.8	274	2	A95252
14	65	12.8	299	2	G98116

15	65	12.8	508	1	UFBSHS
16	65	12.8	1382	2	T01789
17	64.5	12.7	843	2	S78372
18	64.5	12.7	1002	2	S54252
19	64.5	12.7	1033	2	AT2359
20	64.5	12.1	1036	2	T30839
21	64.5	12.1	1127	2	T20870
22	64	12.6	1561	2	T00248
23	63.5	12.5	220	2	E70749
24	63.5	12.5	240	1	H64457
25	63	12.4	186	2	T07977
26	63	12.4	256	2	D87298
27	63	12.4	350	2	D71817
28	63	11.8	1092	2	JK0312
29	62.5	12.3	2048	2	C84609
30	62	11.6	259	2	T46881
31	62	11.6	297	2	H72870
32	62	11.6	489	2	H64224
33	62	11.6	582	2	T32903
34	62	12.2	770	2	H84463
35	61.5	12.1	296	2	E83292
36	61.5	11.5	563	2	H95212
37	61.5	11.5	563	2	A98077
38	61.5	11.5	799	2	S50773
39	61.5	12.1	1422	2	T24212
40	61	12.0	434	2	S53951
41	61	12.0	724	2	T06668
42	61	11.4	1036	2	T05687
43	61	12.0	1076	2	S50536
44	61	12.0	2167	2	S50658
45	60.5	11.3	279	2	C81412

## ALIGNMENTS

### RESULT 1

S37310  
protoporphyrin IX magnesium chelatase (EC 4.99.1.-) - garden snapdragon  
C:Species: Antirrhinum majus (garden snapdragon)  
C:Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 08-Oct-1999  
C:Accession: S37310; S35411  
R:Hudson, A.; Carpenter, R.; Doyle, S.; Coen, E.S.  
EMBO J. 12, 3711-3719, 1993  
A:Title: Olive: a key gene required for chlorophyll biosynthesis in Antirrhinum majus  
A:Reference number: S37310; MUID:94008977; PMID:8404842  
A:Accession: S37310  
A:Status: nucleic acid sequence not shown  
A:Molecule type: DNA  
A:Residues: 1-1379 <HUD>  
C:Cross-references: EMBL:X73144; NID:9312128; PIDN:CAA51664.1; PID:g312129  
C:Superfamily: Rhodobacter capsulatus magnesium-protoporphyrin O-methyltransferase  
C:Keywords: chlorophyll biosynthesis; lyase

Alignment Scores:  
Pred. No.: 4.43 Length: 1379  
Score: 69.00 Matches: 24  
Percent Similarity: 40.96% Conservative: 10  
Best Local Similarity: 28.92% Mismatches: 25  
Query Match: 13.61% Indels: 24  
DB: 2 Gaps: 3

US-09-856-221-3 (1-291) x S37310 (1-1379)

QY	99	GCAGGCAAGATGATGCTGGAATTCAT-----	125
DB	1102	AlaValLysMetValAlaGluLeuAspGluProValGluGlnAsnPheValArgLysHis	1121
QY	126	-----CCTGATGAAGTCGCTTAAAGAACCGTCATGGTCGACCAATCTCGT	173
DB	1122	AlaLeuGluGlnAlaLysGluLeuGlyVal---GluValArgGluAlaAlaSerArgIle	1140
QY	174	TTTCACAGCTTCAGTACGCTGTATCAGCAAGATGTTAAC-----	212

histidine ammonia-  
protoporphyrin IX  
DNA-directed RNA p  
deep orange protei  
hypothetical prote  
sarco/endoplasmic  
hypothetical prote  
zinc finger protei  
probable hycp prot  
uridylylate kinase h  
protoporphyrin IX  
cytochrome-c perox  
differentiation-st  
hypothetical prote  
succinate dehydrog  
hypothetical prote  
hypothetical prote  
hypothetical prote  
conserved hypothet  
ABC transporter, p  
hypothetical prote  
probable membrane  
hypothetical prote  
prefoldin A-sensic  
beta-galactosidase  
hypothetical prote  
bud emergence prot  
NOL1/NOP2/sun fam1









QY 138 CCACCTTCATCAGCAATTCAGCAATCATCTTGCCTGCATTTTAATACCTGCAAGC 79  
 Db 85 -----MetasnlleasnglyserasThrLeualacysillecysanillellyslysas 102  
 QY 78 TAAT 75  
 Db 102 pAsn 103

RESULT 13  
 A95252  
 SPFH domain/Band 7 family [imported] - Streptococcus pneumoniae (strain TIGR4)  
 C:Species: Streptococcus pneumoniae  
 C>Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 03-Aug-2001  
 C:Accession: A95252  
 R:Jettellin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid  
 on, J.D.; Unayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapfel,  
 nson, T.; Hickey, E.K.; Holt, I.E.  
 Science 293, 498-506, 2001  
 A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,  
 A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.  
 A:Reference number: A95000; MUID:21357209; PMID:11463916  
 A:Accession: A95252  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-274 <KUR>  
 A:Cross-references: GB:AE005672; PIDN:AAK76210.1; PID:g14973667; GSPDB:GN00164; TIGR:SP4  
 C:Genetics:  
 A:Gene: SP2156

Alignment Scores:  
 Pred. No.: 12.8 Length: 274  
 Score: 65.00 Matches: 25  
 Percent Similarity: 45.56% Conservative: 16  
 Best Local Similarity: 27.78% Mismatches: 39  
 Query Match: 12.82% Indels: 10  
 DB: 2 Gaps: 2

US-09-856-221-3 (1-291) x A95252 (1-274)

QY 36 AGAAGCAGTGGCAGAAATGCTGCAAACTCAAGCGAGTGAATAGCTTTGCAGAGTATTAA 95  
 Db 151 GlnSerMetAsnGluIleAsnAlaGlnArgLysArgValAlaAlaGlnGluLeuAla 170  
 QY 96 AATGACGACGATGATGCTGAAATGATGCTGATGAGTGGCGCTTAGGAAAGCCGT 155  
 Db 171 GluAlaAspLysIleLysIleValThrAlaAlaGluAlaGluAlaGluLysAspArgLeu 190  
 QY 156 CATGGT-----GCACAATCTCGT-----TTTGACAGCTTC 185  
 Db 191 HisGlyValGlyIleAlaGlnGlnArgLysAlaIleValAspGlyLeuAlaGluSerIle 210  
 QY 186 AGTACGCTGTATGACGAAGATGTTAACTCCGGTGAACCAAGCGATGGATCTGTATCTC 245  
 Db 211 ThrGluLeuLysGluAlaAsnValGlyMetThrGluGluGlnIleMetSerIleLeuLeu 230  
 QY 246 TCTTCATCGGTATTGACGACGACGATGAC 275  
 Db 231 ThrAsnGlnTyrLeuAspThrLeuAsnThr 240

RESULT 14  
 G98116  
 Conserved hypothetical protein spr1962 [imported] - Streptococcus pneumoniae (strain R6)  
 C:Species: Streptococcus pneumoniae  
 C>Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 22-Oct-2001  
 C:Accession: G98116  
 R:Hoskins, J.A.; Albom Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; B  
 e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M  
 y, P.; Sun, P.M.; Winkler, M.E.  
 J. Bacteriol. 183, 5709-5717, 2001  
 A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;  
 A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.

A:Reference number: A97872; MUID:21429245; PMID:11544234  
 A:Accession: G98116  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-299 <KUR>  
 A:Cross-references: GB:AE007317; PIDN:AAU00764.1; PID:g15459662; GSPDB:GN00174  
 C:Genetics:  
 A:Gene: spr1962

Alignment Scores:  
 Pred. No.: 12.9 Length: 299  
 Score: 65.00 Matches: 25  
 Percent Similarity: 45.56% Conservative: 16  
 Best Local Similarity: 27.78% Mismatches: 39  
 Query Match: 12.82% Indels: 10  
 DB: 2 Gaps: 2

US-09-856-221-3 (1-291) x G98116 (1-299)

QY 36 AGAAGCAGTGGCAGAAATGCTGCAAACTCAAGCGAGTGAATAGCTTTGCAGAGTATTAA 95  
 Db 176 GlnSerMetAsnGluIleAsnAlaGlnArgLysArgValAlaAlaGlnGluLeuAla 195  
 QY 96 AATGACGACGATGATGCTGAAATGATGCTGATGAGTGGCGCTTAGGAAAGCCGT 155  
 Db 196 GluAlaAspLysIleLysIleValThrAlaAlaGluAlaGluAlaGluLysAspArgLeu 215  
 QY 156 CATGGT-----GCACAATCTCGT-----TTTGACAGCTTC 185  
 Db 216 HisGlyValGlyIleAlaGlnGlnArgLysAlaIleValAspGlyLeuAlaGluSerIle 235  
 QY 186 AGTACGCTGTATGACGAAGATGTTAACTCCGGTGAACCAAGCGATGGATCTGTATCTC 245  
 Db 236 ThrGluLeuLysGluAlaAsnValGlyMetThrGluGluGlnIleMetSerIleLeuLeu 255  
 QY 246 TCTTCATCGGTATTGACGACGACGATGAC 275  
 Db 256 ThrAsnGlnTyrLeuAspThrLeuAsnThr 265

RESULT 15  
 UPBSHS  
 histidine ammonia-lyase (EC 4.3.1.3) huth [similarity] - Bacillus subtilis  
 N:Alternate names: histidase huth  
 C:Species: Bacillus subtilis  
 C>Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 21-Jul-2000  
 C:Accession: S18810; C69643; T47089  
 R:Oda, M.; Sugishita, A.; Furukawa, K.  
 J. Bacteriol. 170, 3199-3205, 1988  
 A:Title: Cloning and nucleotide sequences of histidase and regulatory genes in the Ba  
 A:Reference number: S18808; MUID:88257040; PMID:2454913  
 A:Accession: S18810  
 A:Molecule type: DNA  
 A:Residues: 1-508 <ODA>  
 A:Cross-references: EMBL:M20659; NID:g143074; PIDN:AAA22538.1; PID:g143076  
 A:Experimental source: strain 1A270  
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber  
 C.; Bron, S.; Broutillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;  
 A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,  
 Nature 390, 249-256, 1997  
 A:Authors: Foulger, D.; Fritz, C.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal  
 iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M  
 Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino  
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau  
 Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete  
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanl  
 A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se  
 akouchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya  
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida  
 A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.  
 A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtili  
 A:Reference number: A69580; MUID:98044033; PMID:9384377  
 A:Accession: C69643  
 A:Status: nucleic acid sequence not shown; translation not shown







GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: January 15, 2003, 15:39:33 ; Search time 6.88336 Seconds  
(without alignments)  
3506.895 Million cell updates/sec

Title: US-09-856-221-3  
Perfect score: 507  
Sequence: 1 aatactgtcgaattac.....tacggcctgcatatggtgc 291

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 225784

Minimum DB seq length: 0

Maximum DB seq length: 200000000

post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-O=CGN2.1/USPTO.spool/US09856221/runat\_15012003\_153924\_12445/app\_query.fasta\_1.1948  
-DB=SwissProt\_40 -OFT=fastan -SUFFIX=rs -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09856221 -CGN\_1.1.38 -runat\_15012003\_153924\_12445 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEPOT=120  
-WARN\_TIMEOUT=30 -THREADS=10 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	66.5	12.4	520	1 APN2_YEAST	P38207 saccharomyc
c 2	66.5	12.4	585	1 HMD2_GOSHI	O64966 gossypium h
c 3	65	12.1	252	1 DHSEB_SCHPO	P21911 schizosacch
c 4	65	12.8	508	1 HUTH_BACSI	P10944 bacillus su
5	64.5	12.7	843	1 RPOC_ODOSI	P49467 odontella s
6	64.5	12.7	1002	1 DOR_DROME	Q24314 drosophila
7	63.5	12.5	220	1 Y085_MYCTU	Q10882 mycobacteri
8	63.5	12.5	240	1 PYRH_METJA	Q58656 methanococc
c 9	63	11.8	1092	1 LIFR_MOUSE	P42703 mus musculu
c 10	62	11.6	259	1 DHSEB_PARDE	Q59662 paracoccus
c 11	62	11.6	499	1 Y225_MYCGE	P47467 mycoplasma
c 12	61.5	11.5	799	1 YJV2_YEAST	P40897 saccharomyc
13	61	12.0	434	1 YMS4_YEAST	Q05131 saccharomyc
14	61	12.0	1076	1 YEM3_YEAST	P40021 saccharomyc
15	61	12.0	2167	1 BEM2_YEAST	P39960 saccharomyc
c 16	60.5	11.3	1885	1 RRPO_ACLSA	P54891 apple chlor
17	59.5	11.7	254	1 PDHR_ECOLI	P06957 escherichia
c 18	59.5	11.1	476	1 COX1_PLABE	O99252 plasmidium

c 19	59.5	11.1	628	1 HMD2_GOSHI	O64967 gossypium h
20	59.5	11.7	1007	1 CHC2_HUMAN	Q9Y514 homo sapien
c 21	59	11.6	67	1 HMT2_METTH	O27731 methanobact
c 22	59	11.0	75	1 DEF1_CAPAN	Q43413 capsicum an
c 23	59	11.6	276	1 Y009_METJA	O60320 methanococc
c 24	59	11.0	732	1 YMM1_CAEEL	P34489 caenorhabdi
c 25	59	11.6	870	1 CSX2_SCHPO	O8ue2 schizosacch
26	59	11.6	1101	1 DIA2_HUMAN	O80879 homo sapien
c 27	59	11.6	1313	1 VGLW_PTPV	P03517 punta coro
c 28	58.5	10.9	169	1 CS60_HUMAN	Q99643 h succinate
c 29	58.5	10.9	298	1 DHSEB_CAEEL	Q99545 caenorhabdi
c 30	58.5	10.9	786	1 AAC_ACTUT	P29958 actinoplane
c 31	58.5	11.5	2198	1 YLJ2_CAEEL	P34367 caenorhabdi
c 32	58	10.8	265	1 PYRF_CANGA	P33283 candida gla
c 33	58	11.4	552	1 VNS1_BTIV1	P35932 bluetonque
c 34	58	11.4	552	1 VNS1_BTIV1	P14245 bluetonque
c 35	58	10.8	716	1 RRP2_IAZII	P13175 influenza a
c 36	58	11.4	3054	1 FOLG_TEV	O84517 t genome po
c 37	58	10.8	5376	1 ZAN_MOUSE	O88799 mus musculu
c 38	57.5	11.3	326	1 OMPH_PHOPR	P29739 photobacter
c 39	57.5	11.3	552	1 VNS1_BTIV2	P35931 bluetonque
c 40	57.5	11.3	575	1 FLA2_CAMJE	P22251 campylobact
c 41	57.5	11.3	575	1 FLB2_CAMJE	P22252 campylobact
c 42	57.5	10.7	746	1 GVP7_YEAST	P48365 saccharomyc
c 43	57.5	10.7	756	1 FSP1_HUMAN	Q92674 homo sapien
c 44	57.5	11.3	1649	1 YG44_SCHPO	O60179 schizosacch
c 45	57.5	11.3	1727	1 ALM1_SCHPO	Q90tk5 schizosacch

## ALIGNMENTS

RESULT 1	APN2_YEAST	STANDARD;	PRT;	520 AA.
ID	APN2_YEAST	STANDARD;	PRT;	520 AA.
AC	P38207;			
DT	01-OCT-1994 (Rel. 30, Created)			
DT	01-OCT-1994 (Rel. 30, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	DNA (apurinic or apyrimidinic site) lyase 2 (EC 4.2.99.18) (AP endonuclease 2) (apurinic-apyrimidinic endonuclease 2).			
DE	GN APN2 OR ETH1 OR YBL019W OR YBL0443.			
OS	Saccharomyces cerevisiae (Baker's yeast).			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
OC	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.			
OX	NCBI_TaxID=4932;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=S288c;			
RA	Goffeau A., Jonniaux J.-L., Purnelle B., Skala J., de Wergifosse P.,			
RA	van Dyck L.;			
RL	Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.			
RN	[2]			
RP	CHARACTERIZATION.			
RX	MEDLINE=20357371; PubMed=9765213;			
RA	Johnson R.E., Torres-Ramos C.A., Izumi T., Mitra S., Prakash S.,			
RA	Prakash L.;			
RT	"Identification of APN2, the Saccharomyces cerevisiae homolog of the			
RT	major human Ap endonuclease HAP1, and its role in the repair of			
RT	basal sites.";			
RL	Genes Dev. 12:3137-3143(1998).			
RN	[3]			
RP	CHARACTERIZATION.			
RX	MEDLINE=20357371; PubMed=10806210;			
RA	Unk I., Haracska L., Johnson R.E., Prakash S., Prakash L.;			
RT	"Apurinic endonuclease activity of yeast Apn2 protein.";			
RL	J. Biol. Chem. 275:22427-22434(2000).			
CC	!- FUNCTION: DNA REPAIR ENZYME THAT HYDROLYZES APURINIC/APYRIMIDINIC			
CC	(AP) SITES AND REMOVES 3'-BLOCKING GROUPS PRESENT AT SINGLE STRAND			
CC	BREAKS OF DAMAGED DNA.			
CC	!- CATALYTIC ACTIVITY: The C-O-P bond 3' to the apurinic or			
CC	apyrimidinic site in DNA is broken by a beta-elimination reaction,			
CC	leaving a 3'-terminal unsaturated sugar and a product with a			
CC	terminal 5'-phosphate.			

```

CC CC -!- SUBCELLULAR LOCATION: Nuclear.
CC CC -!- SIMILARITY: BELONGS TO THE AP/EXO FAMILY OF DNA REPAIR ENZYMES.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL; Z35780; CA48438.1; -
CC CC PIR; S45753; S45753.
CC CC HSP; P27695; 1B9N.
CC CC SGD; S0000115; APN2.
CC CC InterPro; IPR000097; Apendonclisel.
CC CC Pfam; PF03372; Exo_endo_phos_1.
CC CC PROSITE; PS00726; AP_NUCLEASE_F1_1; FALSE_NEG.
CC CC PROSITE; PS00727; AP_NUCLEASE_F1_2; FALSE_NEG.
CC CC PROSITE; PS00728; AP_NUCLEASE_F1_3; 1.
CC CC DNA repair; Lyase; Nuclear protein.
CC CC KW
CC CC SEQUENCE 520 AA; 59445 MW; E3947C4D904C53FB CRC64;
CC CC -----
CC CC Alignment Scores:
CC CC Pred. No.: 5,46 Length: 520
CC CC Score: 66.50 Matches: 30
CC CC Percent Similarity: 42.45% Conservative: 15
CC CC Best Local Similarity: 28.30% Mismatches: 37
CC CC Query Match: 12.43% Indels: 24
CC CC DB: 4 Gaps: 4
CC CC -----
US-09-856-221-3 (1-291) x APN2_YEAST (1-520)
QY 284 ATGCAGGCGCTACTGCTGCTCAATACCGATGAAGAGATACAGATC----- 234
Db 247 MetGlyGlyThrLysLeuGluAlaGlnTyrArgAspLysAlaIleGlnPheIleAsn 266
QY 233 -----CATCGCTGTGTTTTCACCGGAGTT----- 210
Db 267 ProAspThrProHisArgArgIlePheAsnGlnIleLeuAlaAspSerLeuLeuProAsp 286
QY 209 -----AACATCTTCGTACAGCGCTACTGAAGCTGCAACAGAGA----- 168
Db 287 AlaSerLysArgGlyIleLeuIleAspThrThrArgLeuIleGlnThrArgAsnArgLeu 306
QY 167 ----TTGTGCACCATGACGGCTTTCCTTAAGCGCCACTTCATCAGCATCAATTTTCAGCAAT 111
Db 307 LysMetTyrThrValTyrAsnMetLeuLys---AsnLeuArgProSerAsnTyrGlySer 325
QY 110 CATCTTGCTGCTGATTTAATCTCTGCAAGCTAATTCAGTGCCTTGAGTTTCAGCAAT 51
Db 326 ArgIleAspPheIleLeuVal-SerLeuLysLeuGluArgCysIleLysAlaAlaAsp 345
QY 50 TGTGCGAGTGTCTG 35
Db 345 eLeuProAspIleLeu 350
CC CC -----
CC CC RESULT 2
CC CC HMDL_GOSHI
CC CC ID HMDL_GOSHI STANDARD; PRT; 585 AA.
CC CC AC O64966.
CC CC DT 16-OCT-2001 (Rel. 40, Created)
CC CC DT 16-OCT-2001 (Rel. 40, Last sequence update)
CC CC DT 16-OCT-2001 (Rel. 40, Last annotation update)
CC CC DE 3-hydroxy-3-methylglutaryl-coenzyme A reductase 1 (EC 1.1.1.34) (HMG-
CC CC CoA reductase 1).
CC CC GN HMG1.
CC CC OS Gossypium hirsutum (Upland cotton).
CC CC OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
CC CC OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
CC CC OC eurosids II; Malvales; Malvaceae; Gossypium.
CC CC OX NCBI_TaxID=3635;

```

```

RN RP SEQUENCE FROM N.A.
RC STRAIN-cv. Acala S22;
RA Loquercio L.L., Wilkins T.A.;
RT *Two genomic clones encoding 3-hydroxy-3-methylglutaryl-coenzyme A
RT reductase from cotton (Gossypium hirsutum L.);".
RL (In) Plant Gene Register PCR98-031.
CC CC FUNCTION: CATALYZES THE SYNTHESIS OF MEVALONATE. THE SPECIFIC
CC CC PRECURSOR OF ALL ISOPRENOID COMPOUNDS PRESENT IN PLANTS.
CC CC -!- CATALYTIC ACTIVITY: (R)-mevalonate + CoA + 2 NADP(+) -> (S)-3-
CC CC hydroxy-3-methylglutaryl-CoA + 2 NADPH.
CC CC -!- PATHWAY: MEVALONATE BIOSYNTHESIS, PLANT ISOPRENOID BIOSYNTHESIS.
CC CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. ENDOPLASMIC
CC CC RETICULUM. ALSO IN MITOCHONDRION AND PLASTID MEMBRANES.
CC CC -!- SIMILARITY: BELONGS TO THE HMG-COA REDUCTASE FAMILY.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL; AF038045; AAC05088.1; -
CC CC InterPro; IPR002202; HMG-CoA_red.
CC CC PIR; IPR004554; HMG-CoA_R_NADP.
CC CC Pfam; PF00368; HMG-CoA_red; 1.
CC CC PRINTS; PR00071; HMGCOARDTASE.
CC CC TIGRfams; TIGR00533; HMG-CoA_R_NADP; 1.
CC CC PROSITE; PS00066; HMG_COA_REDUCTASE_1; 1.
CC CC PROSITE; PS00318; HMG_COA_REDUCTASE_2; 1.
CC CC PROSITE; PS01192; HMG_COA_REDUCTASE_3; 1.
CC CC PROSITE; PS00065; HMG_COA_REDUCTASE_4; 1.
CC CC Oxidoreductase; Glycoprotein; Endoplasmic reticulum; Transmembrane;
CC CC Isoprene biosynthesis; NADP; Multigene family.
CC CC FT DOMAIN 1 97 MEMBRANE-BOUND (BY SIMILARITY).
CC CC FT DOMAIN 98 169 LINKER (BY SIMILARITY).
CC CC FT DOMAIN 170 585 CATALYTIC (BY SIMILARITY).
CC CC FT TRANSMEM 38 58 POTENTIAL.
CC CC FT TRANSMEM 77 97 POTENTIAL.
CC CC FT ACT_SITE 264 264 BY SIMILARITY.
CC CC FT ACT_SITE 472 472 BY SIMILARITY.
CC CC FT ACT_SITE 570 570 GENERAL BASE (BY SIMILARITY).
CC CC FT CARBOHYD 328 328 N-LINKED (GLCNAC...) (POTENTIAL).
CC CC FT CARBOHYD 441 441 N-LINKED (GLCNAC...) (POTENTIAL).
CC CC FT CARBOHYD 574 574 N-LINKED (GLCNAC...) (POTENTIAL).
CC CC SQ SEQUENCE 585 AA; 62835 MW; 1350E7A9168EBE42 CRC64;
CC CC -----
CC CC Alignment Scores:
CC CC Pred. No.: 5,51 Length: 585
CC CC Score: 66.50 Matches: 27
CC CC Percent Similarity: 46.34% Conservative: 11
CC CC Best Local Similarity: 32.93% Mismatches: 31
CC CC Query Match: 12.43% Indels: 13
CC CC DB: 4 Gaps: 4
CC CC -----
US-09-856-221-3 (1-291) x HMDL_GOSHI (1-585)
QY 274 GTACTGTGCTGCTCAATACCGATGAAGAGATACAGATCCATCGCTGTG---TTTCA 218
Db 164 ValThrValMetThrGluAspGluGluIleIleArgSerValValCysGlyMetThr 183
QY 217 CGGAGTTAATCTTCGCTCATACACGTA-----CTGAAGCTGTCAAACAGAT 167
Db 184 Pro-----SertyrSerLeuGluSerLysLeuAspCysLysArgAla 198
QY 166 TGTGCACCATGACGGCTTTCCTTAAAGCGCCACTTCATCAGCATCAATTTTCAGCAATC 107
Db 199 AlaAlaIleArgArgGluAlaLeuGlnArgIleThrGlyLysSerLeuSerGlyLeuPro 218
QY 106 TGTGCTGCATTTTAACTACTGCAAGCTAATTCATCGCTTGAGTTTCAGCAATTCG 47
Db 106 TGTGCTGCATTTTAACTACTGCAAGCTAATTCATCGCTTGAGTTTCAGCAATTCG 47

```



DR SubtIList; BG10667; huth.  
DR InterPro; IPRO01106; Phe/His\_NH3lyase.  
DR Pfam; PF00221; PAL; 1.  
DR TIGRFAMS; TIGR01225; huth; 1.  
DR PROSITE; PS00488; PAL\_HISTIDASE; 1.  
KW Lyase; Histidine metabolism; Complete proteome.  
FT SITE 141 143  
FT SITE 141 143  
SQ SEQUENCE 508 AA; 55674 MW; 869C3239FCC318E0 CRC64; 5-ONE (BY SIMILARITY).

Alignment Scores:  
Pred. No.: 8.3 Length: 508  
Score: 65.00 Matches: 22  
Percent Similarity: 51.35% Conservative: 16  
Best Local Similarity: 29.73% Mismatches: 18  
Query Match: 12.82% Indels: 18  
DB: 1 Gaps: 4

US-09-856-221-3 (1-291) x HUTH\_BACSU (1-508)

QY 108 ATGATGCTGGAATTGATGCTGATGAAGTGGCGTTAAGGAAAGCGGT----- 155  
Db 204 ValValAlaTyIleGluAlaGluLysLeuAlaTyGlnThrGluArgIleAlaSerLeu 223  
QY 156 -----CATGGTGCACATTCGTTTTCGACGCTTCAGTACGCTGTATGACGAAGATGTT 209  
Db 224 ThrIleGluGlyLeuGlnGlyIleIleAspAlaPhe-----AspGluAspIle 239  
QY 210 AAC-----TCGGTGAAGAAAACAGCATGGATCTG-----TATCTC 245  
Db 240 HisLeuAlaArgGlyTyrrGlnGluGlnIleAspValAlaGluArgIleArgPheTyrrLeu 259  
QY 246 TCTTTCATCGGTATTGACGACGACGACGACGATGCGCTGCATATG 287  
Db 260 SerAspSerGlyLeuThrThrSerGlnGlyGlyLeuArgVal 273

RESULT 5  
RPOC\_ODOSI  
ID RPOC\_ODOSI STANDARD; PRT; 843 AA.  
AC P49467;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE DNA-directed RNA polymerase beta' chain (EC 2.7.7.6).  
OS RPOC1.  
OS Odontella sinensis (Marine centric diatom).  
OC Chloroplast.  
OC Eukaryota; stramenopiles; Bacillariophyta; Coscinodiscophyceae;  
OC Bidulphiphyceidae; Eupodiscales; Eupodisaceae; Odontella.  
OX NCBI\_TaxID=2839;  
RN [1]  
SEQUENCE FROM N.A.  
RA Kowalik K.V., Stoebe B., Schaffran I., Kroth-Pancic P., Freier U.;  
RP "The chloroplast genome of a chlorophyll a+c-containing alga,  
RT Odontella sinensis".  
RL Plant Mol. Biol. Rep. 13:336-342(1995).  
CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION  
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS  
CC SUBSTRATES.  
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
CC [RNA][N].  
CC -1- SUBUNIT: IN CHLOROPLAST THE RNA POLYMERASE IS COMPOSED OF FOUR  
CC SUBUNITS: ALPHA, BETA, BETA', AND BETA".  
CC -1- SUBCELLULAR LOCATION: Chloroplast.  
CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```
CC -----
DR EMBL; Z67753; CAA91745.1; -.
DR HSSP; Q9RW06; JHQM.
DR InterPro; IPR000722; RNA_pol.A.
DR Pfam; PF00623; RNA_pol.A; 1.
DR KX Transferase; Transcription; DNA-directed RNA polymerase; Chloroplast.
SQ SEQUENCE 843 AA; 98521 MW; DF31C1527EC5E58E CRC64;

Alignment Scores:
Pred. No.: 9,97 Length: 843
Score: 64.50 Matches: 18
Percent Similarity: 46.67% Conservative: 24
Best Local Similarity: 20.00% Mismatches: 45
Query Match: 12.72% Indels: 3
DB: 1 Gaps: 2

US-09-856-221-3 (1-291) x RPOC_ODOSI (1-843)

QY 22 CAACGGCAGTGCAGCAGCAGTGCAGCAATGCTGCAAACTCAAGCAGCTGAATTAGCT 81
Db 391 GluLeuGluAspLeuAspLeuLeuGluLeuGlnArgThrArgLysPheLeuValIle 410
QY 82 TTGCAGAGTATTAAATGCAAGCAGATGATTGCTGAAATGATG-----CTGATGAAG 135
Db 411 CysSerLysIleLeuHisLysGluLysProIleTyrHisPheLeuArgTrpPheArgLys 430
QY 136 TGGCGCTTAAGGAAGCCGTCATGCTGCACAACTCTCTTTTGACAGCTTCAGTACGCTGT 195
Db 431 TrpGluLeuGlnArgAlaTyrLysLeuArgAspGlnAlaIle---LysArgIleArgIle 449
QY 196 ATGACGAGATGTTAACTCCGGTGAACAAAGCAGATGATCTCTCTCTCTTCATCGG 255
Db 450 LeuGluAsnLeuLeuAlaThrGlySerAsnProAlaTrpMetIleLeuThrIleLeuPro 469
QY 256 TATTGACACACAGCAGTACGCGCCCTGCAT 285
Db 470 VallieProProAlaLeuArgPrometIle 479

RESULT 6
DOR_DROME
ID DOR_DROME STANDARD; PRT; 1002 AA.
AC Q24314;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Deep orange protein.
GN DOR OR EGL71E4.1.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97218037; PubMed=9065698;
RA Shestopal S.A., Makum I.V., Belyaeva E.S., Ashburner M.;
RT "Molecular characterization of the deep orange (dor) gene of
RL Drosophila melanogaster."
RN [2]
RP Mol. Genet. 253:642-648(1997).
RC SEQUENCE FROM N.A.
RX STRAIN=Oregon-R.
RX MEDLINE=20196011; PubMed=10731137;
RA Benos P.V., Gatt M.K., Ashburner M., Murphy L., Harris D.,
RA Barrell B.G., Ferraz C., Vidal S., Brun C., Demallies J., Cadieu E.,
RA Dreano S., Gloux S., Lelaure V., Mottier S., Galibert F., Borkova D.,
RA Minana B., Kafatos F.C., Louis C., Siden-Kiamos I., Bolshakov S.,
RA Papagiannakis G., Spanos L., Cox S., Madueno E., de Pablos B.,
RA Modolell J., Peter A., Schoettler P., Werner M., Mourikioti F.,
RA Beinert N., Dowe G., Schaefer U., Jaecle H., Bucheton A.,
RA Callister D.M., Campbell L.A., Darlamitsou A., Henderson N.S.,
RA McMillan P.J., Salles C., Tait E.A., Valenti P., Saunders R.D.C.,
RA Glover D.M.;
```

```
RT -----
RL Science 287:2220-2222(2000).
CC -1- SIMILARITY: SOME, TO YEAST PEP3.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X86683; CAA60382.1; -.
DR EMBL; AL021726; CAA16809.1; -.
DR FlyBase; FBgn0000482; dor.
DR InterPro; IPR000547; Clathrin_repeat.
DR InterPro; IPR001841; Znf_ring.
DR SMART; SM00299; CLH; 1.
DR SMART; SM00184; RING; 1.
KW Zinc-finger; Transmembrane.
FT ZN_FING 885 910 C3H2C-TYPE.
FT TRANSMEM 971 991 POTENTIAL.
FT CONFLICT 169 169 A -> P (IN REF. 1).
FT CONFLICT 581 581 Q -> H (IN REF. 1).
FT CONFLICT 865 865 A -> V (IN REF. 1).
SQ SEQUENCE 1002 AA; 115305 MW; D59690A0FC95182F CRC64;

Alignment Scores:
Pred. No.: 10.1 Length: 1002
Score: 64.50 Matches: 22
Percent Similarity: 42.16% Conservative: 21
Best Local Similarity: 21.57% Mismatches: 36
Query Match: 12.72% Indels: 23
DB: 1 Gaps: 3

US-09-856-221-3 (1-291) x DOR_DROME (1-1002)

QY 1 AATACCTTGCTCAACTACTGACGCGCAGGATGCAGAGCACTGGCAGAAATGTGTCGAA 60
Db 698 AsnPheLeuLeuHisLeuTyrAlaGluHisGluProLysLeuLeuMetLysTyrLeuGlu 717
QY 61 ACTCAAGCAGTGAATTAGCTTTGCAGAGTATTAAATG----- 99
Db 718 IlegInGlyArgAspGluSerLeuValHisTyrAspIleTyrTyrAlaHisLysValCys 737
QY 100 -----CAGGCAAGATGATGCTGAAATGATGCTGAAGTGAAGTGGCGC 141
Db 738 ThrAspLeuAspValLysGluAlaArgValPheLeuGluCysMetLeuArgLysTrp--- 756
QY 142 TTAAGGAAAGCCGTCATGTCGACAACTCTGTTTGACAGCTTCAGTACGCTGTATGAGC 201
Db 757 -----IleSerAlaValAspLeuAlaLeuThrPheAspMetLysLeuAlaLys 772
QY 202 AAGATGTTA-----ACTCCGCTGAAACAAACAGCAGTGGATCTGTATCTCT 246
Db 773 GluThrAlaSerArgProSerAspSerLysIleArgLysLeuTrpLeuArgIleAla 792
QY 247 CTTCAT 252
Db 793 TyrHis 794

RESULT 7
Y085_MYCTU
ID Y085_MYCTU STANDARD; PRT; 220 AA.
AC Q10882;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein rv0085.
GN RV0085 OR MT0092 OR MTCY251.03.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
```

OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1773;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H37RV;  
 RX MEDLINE=98295987; PubMed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;  
 RA "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence.";  
 RL Nature 393:537-544 (1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CDC 1551 / Oshkosh;  
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 RA Peterson J., Deboy R., Dodson R., Gwinn M.L., Hart D., Hickey E., S.L.,  
 RA Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
 RA Bishai W.;  
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
 laboratory strains.";  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; 274410; CA98921.1; -;  
 DR EMBL; AF006920; AAK44317.1; -;  
 DR TIGR; MT0092; -;  
 DR Tuberculist; RV0085; -;  
 DR KW Hypothetical protein; Transmembrane; Complete proteome.  
 FT TRANSMEM 6 26 POTENTIAL.  
 FT TRANSMEM 33 53 POTENTIAL.  
 FT TRANSMEM 59 79 POTENTIAL.  
 FT TRANSMEM 103 123 POTENTIAL.  
 FT TRANSMEM 126 146 POTENTIAL.  
 FT TRANSMEM 157 177 POTENTIAL.  
 FT TRANSMEM 179 199 POTENTIAL.  
 SQ SEQUENCE 220 AA; 23042 MW; CB76F31A2972FD3C CRC64;  
 Alignment Scores:  
 Pred. No.: 11.8 Length: 220  
 Score: 63.50 Matches: 30  
 Percent Similarity: 41.41% Conservative: 11  
 Best Local Similarity: 30.30% Mismatches: 29  
 Query Match: 12.52% Indels: 29  
 DB: 1 Gaps: 4  
 US-09-856-221-3 (1-291) x Y085\_MYCTU (1-220)  
 QY 16 ATTACTGACGGCAGGATGCGAGAGCTGCGAGATTCGTCGCAACTCAAGCAGTGA 75  
 Db 24 IIEValTrpArgArgArgArgArgArgArgArgArgArgArgArgArgArgArg 43  
 QY 76 TTAGCT-----TTGCAGAGTATATAAATGCAGCAAGATGATTC----- 114  
 Db 44 LeuAlaAlaIleProLeuLeuArgGlyIleArgAspAsnAspAlaLeuIleAlaVal 63  
 QY 115 -----CTGAATTGATCGATGAGTGGCGCTTTAAGAAAGCCGTC 156  
 Db 64 GlyIleAlaValLeuAlaLeuArgArgArgArgArgArgArgArgArgArgArg 83

QY 157 ATGTGTCAC-----AATCTCGTTTTCAGACGCTTCAGTACGC 192  
 Db 84 GlyAlaGluAlaAlaAlaGlnArgGluAlaThrProLeuValAsnThrAlaSerLeu 103  
 QY 193 TGTATGACGACAGATGTTAACT-----CCGCTGAAAC 225  
 Db 104 LeuIleThrAlaGlyLeuThrLeuThrAlaPheAlaIleThrGlnProValAsn 122  
 RESULT 8  
 PYRH\_METJA  
 ID PYRH\_METJA STANDARD; PRT; 240 AA.  
 AC Q58656;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Putative uridylyate kinase (EC 2.7.4.-) (UK) (Uridine monophosphate  
 DE kinase) (UMP kinase).  
 GN MJ1259.  
 OS Methanococcus jannaschii.  
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;  
 OC Methanocaldococaceae; Methanocaldococcus.  
 OX NCBI\_TaxID=2190;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;  
 RX MEDLINE=96337999; PubMed=8688087;  
 RA Sult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
 RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,  
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,  
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,  
 RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,  
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;  
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus  
 jannaschii.";  
 RL Science 273:1058-1073 (1996).  
 CC -!- FUNCTION: URIDINE MONOPHOSPHATE KINASE (BY SIMILARITY).  
 CC -!- CATALYTIC ACTIVITY: ATP + UMP = ADP + UDP.  
 CC -!- SIMILARITY: TO OTHER UMP KINASE; SOME, TO ASPARTOKINASES AND TO  
 CC GLUTAMATE KINASES.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; U67566; AAB99262.1; -;  
 DR TIGR; MJ1259; -;  
 DR InterPro; IPR001048; Aa\_kinase.  
 DR Pfam; PF00696; aakinase; 1.  
 KW Hypothetical protein; Transferase; Kinase; Pyrimidine biosynthesis;  
 KW Complete proteome.  
 SQ SEQUENCE 240 AA; 25611 MW; E4C49965C071E66F CRC64;  
 Alignment Scores:  
 Pred. No.: 11.9 Length: 240  
 Score: 63.50 Matches: 22  
 Percent Similarity: 49.41% Conservative: 20  
 Best Local Similarity: 25.88% Mismatches: 26  
 Query Match: 12.52% Indels: 17  
 DB: 1 Gaps: 3  
 US-09-856-221-3 (1-291) x PYRH\_METJA (1-240)  
 QY 39 ACCACTGCGACATTCCTCGCAACTCAAGCAGTGAATTACCTTCGACAGTATTAAT 98  
 Db 135 ThrThrAspAlaValAlaAla-----SerLeuAlaGluPheIleAsn 148



QY 99 GCAGCAGATGATGCTGAAATTCATGCTGATGAAGTGGCGCTTAAGAA---AGCGGT 155  
 Db 149 AlaAspLeuValIleGlyThrAsnValAspGlyValThrAspLysAspProAsnLys 168  
 QY 156 CATGTCGCACAACTCGTTTTCACAGCTTCAGTACGCTGATGACGAAATGTTAACTCC 215  
 Db 169 TyrGluAspAlaLysLysPheAspLysMetSerAla----- 180  
 QY 216 GGTGAAACAAACGATGATGCTGATCTCTTCATCGTATGACGACACGAGTACG 275  
 Db 181 -----LysGluLeuValAspLeuAlaIleSerSerLeuLysAlaGlySerSer 198  
 QY 276 GCCTCGCATGGTG 290  
 Db 199 ValValAspLeuLeu 203  
 RESULT 9  
 LIFR\_MOUSE STANDARD; PRT: 1092 AA.  
 AC P42703;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Leukemia inhibitory factor receptor precursor (LIF-R) (D-factor/LIF  
 DE receptor).  
 GN LIFR.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 [1]  
 RN MEDLINE=92007727; PubMed=1915266;  
 RA Gearing D.P., Phut C.J., Vandenbos T., Gimpel S.D., Delaney P.B.,  
 RA King J., Price V., Cosman D., Beckmann M.P.;  
 RA "Leukemia inhibitory factor receptor is structurally related to the  
 RT IL-6 signal transducer, gp130.";  
 RL EMBO J. 10:2839-2848(1991).  
 [2]  
 RN SEQUENCE FROM N.A. (ISOFORM 2).  
 RC STRAIN=ICR; TISSUE=Liver;  
 RX MEDLINE=94039833; PubMed=7901054;  
 RA Tomida M., Yamamoto-Yamaguchi Y., Hozumi M.;  
 RT "Pregnancy associated increase in mRNA for soluble D-factor/LIF  
 RL FEBS Lett. 334:193-197(1993).  
 [3]  
 RN SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
 RX MEDLINE=94334302; PubMed=8056772;  
 RA Tomida M., Yamamoto-Yamaguchi Y., Hozumi M.;  
 RT "Three different cDNAs encoding mouse D-factor/LIF receptor.";  
 RL J. Biochem. 115:557-562(1994).  
 CC -1- FUNCTION: SIGNAL-TRANSDUCING MOLECULE. MAY HAVE A COMMON PATHWAY  
 CC WITH GP130. THE SOLUBLE FORM INHIBITS THE BIOLOGICAL ACTIVITY OF  
 CC LIF BY BLOCKING ITS BINDING TO RECEPTORS ON TARGET CELLS.  
 CC -1- SUBUNIT: HETERODIMER COMPOSED OF LIFR AND GP130.  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. THERE IS A  
 CC MEMBRANE-BOUND AND A SECRETED FORM.  
 CC -1- ALTERNATIVE PRODUCTS: 2 isoforms; 1/membrane (shown here) and  
 CC 2/secreted; may be produced by alternative splicing.  
 CC -1- TISSUE SPECIFICITY: PLACENTA, LIVER, KIDNEY, HEART, LUNG, BRAIN,  
 CC AND EMBRYOS. THE LIVER MAY BE THE PRIMARY SITE OF SYNTHESIS OF  
 CC THE SECRETED FORM.  
 CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 4 FIBRONECTIN TYPE III-LIKE DOMAINS.  
 CC -1- SIMILARITY: BELONGS TO THE CYTOKINE FAMILY OF RECEPTORS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
 CC  
 CC  
 DR EMBL; S73496; AAC60698.1; -  
 DR EMBL; S73495; AAC60697.1; -  
 DR EMBL; D26177; BAA05165.1; -  
 DR EMBL; D17444; BAA04258.1; -  
 DR MGI; 96788; Lifer; CRLA.  
 DR InterPro; IPR002996; FN.III.  
 DR InterPro; IPR003961; FN.III.  
 DR Pfam; PF00041; fn3; 4.  
 DR SMART; SM00060; FN3; 3.  
 DR PROSITE; PS01353; HEMATOPO\_REC\_L\_F2; 1.  
 KW Receptor; Transmembrane; Glycoprotein; Immunoglobulin domain; Signal;  
 KW Alternative splicing; Repeat.  
 FT SIGNAL 1 43 POTENTIAL.  
 FT CHAIN 44 1092 LEUKEMIA INHIBITORY FACTOR RECEPTOR.  
 FT DOMAIN 44 828 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 829 853 POTENTIAL.  
 FT DOMAIN 854 1092 CYTOPLASMIC (POTENTIAL).  
 FT DISULFID 53 63 BY SIMILARITY.  
 FT CARBOHYD 164 164 N-LINKED (GLCNAC. -) (POTENTIAL).  
 FT CARBOHYD 199 199 N-LINKED (GLCNAC. -) (POTENTIAL).  
 FT CARBOHYD 238 238 N-LINKED (GLCNAC. -) (POTENTIAL).  
 FT CARBOHYD 261 261 N-LINKED (GLCNAC. -) (POTENTIAL).  
 FT CARBOHYD 385 385 N-LINKED (GLCNAC. -) (POTENTIAL).  
 FT CARBOHYD 402 402 N-LINKED (GLCNAC. -) (POTENTIAL).  
 FT CARBOHYD 421 421 N-LINKED (GLCNAC. -) (POTENTIAL).  
 FT CARBOHYD 440 440 N-LINKED (GLCNAC. -) (POTENTIAL).  
 FT CARBOHYD 453 453 N-LINKED (GLCNAC. -) (POTENTIAL).  
 FT CARBOHYD 476 476 N-LINKED (GLCNAC. -) (POTENTIAL).  
 FT CARBOHYD 567 567 N-LINKED (GLCNAC. -) (POTENTIAL).  
 FT CARBOHYD 647 647 N-LINKED (GLCNAC. -) (POTENTIAL).  
 FT CARBOHYD 658 658 N-LINKED (GLCNAC. -) (POTENTIAL).  
 FT CARBOHYD 675 675 N-LINKED (GLCNAC. -) (POTENTIAL).  
 FT CARBOHYD 724 724 N-LINKED (GLCNAC. -) (POTENTIAL).  
 FT CARBOHYD 782 782 N-LINKED (GLCNAC. -) (POTENTIAL).  
 FT VARSPLIC 718 719 AP -> EA (IN ISOFORM 2).  
 FT VARSPLIC 720 1092 MISSING (IN ISOFORM 2).  
 SQ SEQUENCE 1092 AA; 122573 MW; 6F02BEC8E154DE70 CRC64;  
 Alignment Scores:  
 Pred. No.: 15.5 Length: 1092  
 Score: 63.00 Matches: 27  
 Percent Similarity: 38.78% Conservative: 11  
 Best Local Similarity: 27.55% Mismatches: 30  
 Query Match: 11.78% Indels: 30  
 DE: 1 Gaps: 5  
 US-09-856-221-3 (1-291) x LIFR\_MOUSE (1-1092)  
 QY 290 CACCATATGCGAGGCGGTACT-----GCTGGTCTCAATACCGATGAAGA 246  
 Db 378 HisArgIleGluGlyLeuThrAsnGluThrTyrArgLeuGlyValGlnMethisProgly 397  
 QY 245 GAGATACAGATCCATCGCTGTTTTCACCGAGCTTAAC----- 207  
 Db 398 Gln---GluIleHisAsnPheThrLeuThrGlyArgAsnProLeuGlyGlnAlaGlnSer 416  
 QY 206 ATCTTCGTCATACAGCGTACTGAAGCTGTC----- 177  
 Db 417 AlaValIleAsnValThrGluArgValAlaProHisAspProThrSerLeuLysVal 436  
 QY 176 -----AAACGAGATTGTGCACCATGACGCGTTTCCTTAAGCGCCACTTCATCAGC 126  
 Db 437 LysAspIleAsnSerThrValValThrPheSerTrpTyrLeuProGlyAsnPheThrLys 456  
 QY 125 ATCAATTTACGACATCATCTCGCTGCAATTTTAATATCTCTGCAAGCTAATTA 72  
 Db 457 IleAsn-----LeuLeuCysGlnIleGluIleCysLysAlaAsnSer 470  
 RESULT 10  
 DHSB\_PARDE

```
ID DSB_PARDE STANDARD; PRT; 259 AA.
AC Q59662;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Succinate dehydrogenase iron-sulfur protein (EC 1.3.99.1).
GN SDHB.
OS Paracoccus denitrificans.
OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
OC Paracoccus.
OX NCBI_TaxID=256;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=ATCC 13543;
RC Dickinson M.A., Dhawan T., Gunsalus R.P., Schroeder I., Cecchini G.;
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
RL -! CATALYTIC ACTIVITY: Succinate + acceptor + reduced
CC acceptor.
CC -! COFACTOR: BINDS THREE DIFFERENT IRON-SULFUR CLUSTERS: A 2FE-2S,
CC A 3FE-4S AND A 4FE-4S (BY SIMILARITY).
CC -! PATHWAY: Tricarboxylic acid cycle.
CC -! SUBUNIT: PART OF AN ENZYME COMPLEX CONTAINING FOUR SUBUNITS: A
CC FLAVOPROTEIN, AN IRON-SULFUR, CYTOCHROME B-556, AND AN HYDROPHOBIC
CC ANCHOR PROTEIN.
CC -! SIMILARITY: THE IRON-SULFUR CENTERS ARE SIMILAR TO THOSE OF
CC 'PLANT-TYPE' 2FE-2S AND 'BACTERIAL-TYPE' 4FE-4S FERREDOXINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U31902; AAA75178.1;
DR HSP; P00364; LFUM.
DR InterPro; IPR000564; 2Fe2S_ferredoxin.
DR InterPro; IPR001450; 4Fe4S_ferredoxin.
DR InterPro; IPR004489; DhsB.
DR Pfam; PF00111; fer2; 1.
DR TIGRfams; TIGR00384; dhsb; 1.
DR PROSITE; PS00197; 2FE2S_FERREDOXIN; 1.
DR PROSITE; PS00198; 4FE4S_FERREDOXIN; 1.
KW Tricarboxylic acid cycle, Iron-sulfur; Oxidoreductase;
KW Electron transport; 3Fe-4S; 4Fe-4S.
FT METAL 80 85 IRON-SULFUR S1 (2FE-2S) (BY SIMILARITY).
FT METAL 85 85 IRON-SULFUR S1 (2FE-2S) (BY SIMILARITY).
FT METAL 88 88 IRON-SULFUR S1 (2FE-2S) (BY SIMILARITY).
FT METAL 100 100 IRON-SULFUR S1 (2FE-2S) (BY SIMILARITY).
FT METAL 170 170 IRON-SULFUR S2 (4FE-4S) (BY SIMILARITY).
FT METAL 173 173 IRON-SULFUR S2 (4FE-4S) (BY SIMILARITY).
FT METAL 176 176 IRON-SULFUR S2 (4FE-4S) (BY SIMILARITY).
FT METAL 180 180 IRON-SULFUR S3 (3FE-4S) (BY SIMILARITY).
FT METAL 227 227 IRON-SULFUR S3 (3FE-4S) (BY SIMILARITY).
FT METAL 233 233 IRON-SULFUR S2 (4FE-4S) (BY SIMILARITY).
FT METAL 237 237 IRON-SULFUR S2 (4FE-4S) (BY SIMILARITY).
SQ SEQUENCE 259 AA; 29551 MW; F70C21BBA8205B4F CRC64;

Alignment Scores:
Pred. No.: 18.2 Length: 259
Score: 62.00 Matches: 19
Percent Similarity: 47.27% Conservative: 7
Best Local Similarity: 34.55% Mismatches: 12
Query Match: 11.59% Indels: 17
DB: 1 Gaps: 2

US-09-856-221-3 (1-291) x DhsB_PARDE (1-259)
QY 258 ATACCGATGAGAGATACAGATCGCTGTTTTCACCGAGGTAAACATCTCGT 199
DB 64 ILeYsIleLysAsnGluIleAspProThrLeuThrPheArg----- 78
```

```
QY 198 CATACAGCGTACTGAAGCTGTCAAAACAGAGATTGTGCACCATGACGGCTTCTCTTAAGCG 139
DB 79 -----SerCysArgGluGlyIle-CysGlySerCysAla----- 89
QY 138 CCACCTTCATCAGCATCAATTCAGCAATCATCTTGCCTGCATT 96
DB 90 -----MetAsnIleAspGlyGlyAsnHisLeuAlaCysIle 101
RESULT 11
Y225_MYCGE
ID Y225_MYCGE STANDARD; PRT; 489 AA.
AC P47467; O49208;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MG225.
GN MG225
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA Nguyen D.F., Weidman J.F., Small K.V., Sandusky M., Fuhmann J.L.,
RA Tomb J.-F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT "The minimal gene complement of Mycoplasma genitalium."
RL Science 270:397-403(1995).
RN [2]
RP SEQUENCE OF 225-286 FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RA Peterson S.N., Hu P.-C., Bott K.F., Hutchison C.A. III;
RT sequencing.
RT "A survey of the Mycoplasma genitalium genome by using random
CC -! SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -! SIMILARITY: TO M.GENITALIUM MG226.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U39702; AAC71446.1;
DR EMBL; U01756; AAD10570.1;
DR TIGR; MG225;
DR InterPro; IPR002293; AA/rel_pmease1.
DR InterPro; IPR004841; Pemease.
DR Pfam; PF00324; aa_pemeases; 1.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 14 34 POTENTIAL.
FT TRANSMEM 36 56 POTENTIAL.
FT TRANSMEM 100 120 POTENTIAL.
FT TRANSMEM 127 147 POTENTIAL.
FT TRANSMEM 158 178 POTENTIAL.
FT TRANSMEM 203 223 POTENTIAL.
FT TRANSMEM 241 261 POTENTIAL.
FT TRANSMEM 286 306 POTENTIAL.
FT TRANSMEM 344 364 POTENTIAL.
FT TRANSMEM 380 400 POTENTIAL.
FT TRANSMEM 419 439 POTENTIAL.
FT TRANSMEM 449 469 POTENTIAL.
FT CONFLICT 282 286 IFKKA -> TLEDI (IN REF. 2).
SQ SEQUENCE 489 AA; 54766 MW; CAPEE4E83DDF8A9 CRC64;
```





CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC	EMBL; Z35159; CAA84524.1; -	
DR	EMBL; L33832; AAA57132.1; -	
DR	EMBL; U18917; AAB64682.1; -	
DR	SGD; S0000957; BEM2.	
DR	InterPro; IPR001849; PH.	
DR	InterPro; IPR000651; RasGEFN.	
DR	InterPro; IPR001895; RasGRF_CDC25.	
DR	InterPro; IPR000198; RhoGAP.	
DR	Pfam; PF00169; PH; 1.	
DR	Pfam; PF00617; RasGEF; 1.	
DR	Pfam; PF00618; RasGEFN; 1.	
DR	Pfam; PF00620; RhoGAP; 1.	
DR	SMART; SM00233; PH; 1.	
DR	SMART; SM00147; RasGEF; 1.	
DR	SMART; SM00229; RasGEFN; 1.	
DR	SMART; SM00324; RhoGAP; 1.	
DR	PROSITE; PS50003; PH_DOMAIN; 1.	
KW	GTPase activation.	
FT	DOMAIN 1914 1948	
FT	DOMAIN 1981 2134	RHO-GAP.
FT	DOMAIN 16 25	POLY-SER.
FT	DOMAIN 35 43	POLY-SER.
FT	DOMAIN 58 63	POLY-HIS.
FT	DOMAIN 198 208	POLY-ASN.
FT	DOMAIN 253 260	POLY-SER.
FT	DOMAIN 1161 1165	POLY-THR.
SQ	SEQUENCE 2157 AA; 245428 MW; DB7FF4CD417E898F0 CRC64;	

Alignment Scores:		
Pred. No.:	28.8	Length: 2167
Score:	61.00	Matches: 21
Percent Similarity:	49.35%	Conservative: 17
Best Local Similarity:	27.27%	Mismatches: 21
Query Match:	12.03%	Indels: 18
DB:	1	Gaps: 3

```

US-09-856-221-3 (1-291) x BEM2_YEAST (1-2167)

QY 31 GATCAGAAAGCACTGGCAGAAATTCGTGCAAACTCAAGGC-----AGTGAATTA----- 78
      :::::|||||:::|||||
Db 1709 GluSerGluAlaIleSerGluArgPheGlnGluGlnGlyValPheAsnGluIleLeuVal 1728
      :::::|||||:::|||||

QY 79 ---GCTTTGCAGACATATTAATTCAGCAGCAAGCATGATTCGTGAAATTGATGCTGATGAAG 135
      :::::|||||:::|||||
Db 1729 AsnGluIleGluIlyIleLysArgGluAlaArgIlyLeuGluValLeuLeuAspGlnGlu 1748
      :::::|||||:::|||||

QY 136 TGGCGCTTTAAGGAAAGCGGTGATGCTGGCAACATCTCGTTTGCACCTTCAGTCAAGCTGT 195
      |||||:::|||||
Db 1749 LysIleLeuLysAsnSerAlaAlaLeuHisGlnAlaVal-----1761

QY 196 ATGACGAAGATGTTAACTCCGGTGAAACCAAGCGATGGATCTGTATCTCT 246
      ||| |||||:::::: |||||
Db 1762 -----ProLysIlyAsnArgLysSerValIleIleSer 1772

```

Search completed: January 15, 2003, 15:46:31  
Job time : 10.8834 secs



GenCore version 5.1.3  
 Copyright (c) 1993 - 2003 CompuGen Ltd.  
 OM nucleic - protein search, using frame\_plus\_n2p model  
 Run on: January 15, 2003, 15:44:13 ; Search time 27.4148 Seconds  
 (without alignments)  
 4374.264 Million cell updates/sec

Title: US-09-856-221-3  
 Perfect score: 507  
 Sequence: 1 aataactgtcacaattac.....tacggccgtcatatggtgc 291

Scoring table: BLOSUM62  
 Xgapop 10.0 , Xgapext 0.5  
 Ygapop 10.0 , Ygapext 0.5  
 Fgapop 6.0 , Fgapext 7.0  
 Delop 6.0 , Delext 7.0

Searched: 671580 seqs, 206047115 residues  
 Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0  
 Maximum DB seq length: 2000000000  
 Post-processing: Minimum Match 0%  
 Maximum Match 100%  
 Listing first 45 summaries

Command line parameters:  
 -MODEL=frame\_n2p.model -DEV=xlp  
 -DB=SPTRMBL\_21 -QFMT=fastan -SUFFIX=tspt -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
 -UNIT5=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS-human40.cdi -LIST=45  
 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
 -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
 -USER=US09856221.ecgn\_1.171 -runat\_15012003\_153925\_12457 -NCPU=6 -ICPU=3  
 -NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
 -WARN\_TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SPTRMBL\_21:  
 1: sp\_archaea:  
 2: sp\_bacteria:  
 3: sp\_fungi:  
 4: sp\_human:  
 5: sp\_invertebrate:  
 6: sp\_mammal:  
 7: sp\_mhc:  
 8: sp\_organelle:  
 9: sp\_phage:  
 10: sp\_plant:  
 11: sp\_rodent:  
 12: sp\_virus:  
 13: sp\_vertebrate:  
 14: sp\_unclassified:  
 15: sp\_rvirus:  
 16: sp\_bacteriaph:  
 17: sp\_archaeap:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Query Match Length	Description
1	399	78.7	2523 2 Q93RPO Q93rp0 xenorhabdus

2	293	57.8	2376	2	Q9F9Z3	Q9f9z3 serratia en
3	220	43.4	2516	2	Q9RN43	Q9rn43 photorhabdu
4	208	41.0	2504	2	Q85160	Q85160 photorhabdu
5	184	36.3	2538	2	Q93RN7	Q93rn7 xenorhabdus
6	109	21.5	1565	2	Q85156	Q85156 photorhabdu
7	71	14.0	296	16	Q99Y37	Q99y37 streptococc
8	71	14.0	1693	2	Q9F5F7	Q9f5f7 agrobacteri
9	69	13.6	1379	10	Q07893	Q07893 antirrhinum
10	68	13.4	299	5	Q16734	Q16734 caenorhabdi
11	68	13.4	329	16	Q34803	Q34803 bacillus su
12	68	13.4	626	3	Q59966	Q59966 fusarium ox
13	68	13.4	1383	10	Q65808	Q65808 glycine max
14	68	13.4	1693	2	Q9R612	Q9r612 agrobacteri
15	68	13.4	1693	16	Q80634	Q8u634 agrobacteri
16	67	13.2	1189	2	Q85152	Q85152 photorhabdu
c 17	67	12.5	1345	5	Q96236	Q96236 plasmodium
c 18	67	13.2	1380	10	Q40001	Q40001 hordeum vul
19	66	13.0	218	15	Q927Y1	Q927y1 listeria in
c 20	66	12.3	1349	12	Q9DWB6	Q9dwb6 rat cytoleg
21	66	13.0	1463	5	Q9GYZ0	Q9gyz0 strongyloce
c 22	65.5	12.2	930	10	Q22161	Q22161 arabidopsis
23	65	12.8	138	2	Q9X2K5	Q9x2k5 acinetobact
24	65	12.8	274	16	Q97N99	Q97n99 streptococc
25	65	12.8	1279	2	Q93RP4	Q93rp4 xenorhabdus
26	65	12.8	1382	10	Q22435	Q22435 nicotiana t
c 27	64.5	12.1	228	8	Q8WCM5	Q8wcm5 lonicera sp
c 28	64.5	12.1	370	12	Q81967	Q81967 human papil
29	64.5	12.7	1002	5	Q9S677	Q9s677 drosophila
30	64.5	12.7	1033	16	Q8YXX6	Q8yxx6 anabaena sp
c 31	64.5	12.7	1036	5	Q76974	Q76974 paramecium
c 32	64.5	12.1	1037	5	Q9N9D8	Q9n9d8 paramecium
c 33	64.5	12.1	1127	5	Q27493	Q27493 caenorhabdi
c 34	64.5	12.1	464	5	Q9V8D8	Q9v8d8 drosophila
35	64	12.6	782	5	Q61073	Q61073 paramecium
c 36	64	12.0	784	10	Q9XE10	Q9xe10 stevia reba
37	64	12.6	1040	11	Q92518	Q92518 mus musculu
38	64	12.6	1561	11	Q88286	Q88286 mus musculu
39	63.5	12.5	475	16	Q8XYA8	Q8xya8 raietonia s
40	63.5	11.9	578	10	Q9FM59	Q9f5m9 arabidopsis
c 41	63.5	12.4	186	10	Q22590	Q22590 onobrychis
42	63	12.4	256	16	Q9AB37	Q9ab37 caulobacter
43	63	12.4	338	5	Q9VW96	Q9vw96 drosophila
c 44	63	11.8	350	16	Q9ZJF8	Q9zjf8 helicobacte
45	63	12.4				

ALIGNMENTS

RESULT 1	
Q93RPO	PRELIMINARY; PRT; 2523 AA.
AC Q93RPO	01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001	(TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002	(TrEMBLrel. 20, Last annotation update)
DE	Xptal protein.
GN	XPTAL.
OS	Xenorhabdus nematophilus.
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
CC	Xenorhabdus.
OX	NCBI_Taxid=628;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=PMF1296;
RX	MEDLINE=21218513; PubMed=11319082;
RA	Morgan J.A.W., Sergeant M., Ellis D., Ousley M., Jarrett P.J.
RT	"Sequence analysis of insecticidal genes from Xenorhabdus nematophilus
RT	PMF1296.";
RL	Appl. Environ. Microbiol. 67:2062-2069(2001).
DR	EMBL; AJ308438; CAC38401.1; -
DR	InterPro; IPR000953; Chromo.
DR	SMART; SM00298; CHROMO; 1.
SQ	SEQUENCE 2523 AA; 286999 MW; 3159852E0655B5B1 CRC64;

```
Alignment Scores:
Pred. No.: 1,01e-41 Length: 2523
Score: 399.00 Matches: 85
Percent Similarity: 97.92% Conservatives: 9
Best Local Similarity: 88.54% Mismatches: 2
Query Match: 78.70% Indels: 1
DB: 2 Gaps: 0

US-09-856-221-3 (1-291) x Q93RP0 (1-2523)
QY 1 AATACCTTGCCTCAACATTACTGACGCGAGGATGACAGAGCAGCTGGCAGATTGCTGCAA 60
|||||
Db 2054 AnThrLeuLeuSerIleThrGluArgGlnAspAlaGluAlaLeuAlaGluIleLeuGln 2073
|||||
QY 61 ACTCAAGCAGTGAATAGCTTTGTCAGAGTATTAAATGACAGGC-AAGATGATTGCTGAA 119
|||||
Db 2074 ThrGlnGlySerGluLeuAlaLeuGlnSerIleLysMetGlnAspLysValMetAlaGlu 2093
|||||
QY 120 ATTGATGCTGATGAAGTGGCGCTTAAGGAAGCCGTCATGTCACAAATCTGTTTGAC 179
|||||
Db 2094 IleAspAlaAspLysLeuAlaLeuGlnGluSerArgHISGlyAlaGlnSerArgPheAsp 2113
|||||
QY 180 AGCTTCAGTACGCTGTATGACGAAGATGTTAACTCCGCTGAAACAAACGATGATCTG 239
|||||
Db 2114 SerPheAsnThrLeuTyAspGluAspValAsnAlaGlyGluLysGlnAlaMetAspLeu 2133
|||||
QY 240 TATCTCTCTTCATCGGTATTGACACCAGCAGTACGCCCTGCATATG 287
|||||
Db 2134 TyrLeuSerSerValLeuSerThrSerglyThrAlaLeuHisMet 2149

RESULT 2
Q9F9Z3 ID Q9F9Z3 PRELIMINARY; PRT; 2376 AA.
AC Q9F9Z3;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE SEPA.
GN Serratia entomophila.
OS Plasmid pADAP.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Serratia.
OX NCBI_TaxID=42906;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20416224; PubMed=10960097;
RA Hurst M.R., Glare T.R., Jackson T.A., Ronson C.W.;
RT "Plasmid-located pathogenicity determinants of Serratia entomophila,
RT the causal agent of amber disease of grass grub, show similarity to
RT the insecticidal toxins of Photobacterium luminescens.";
RL J. Bacteriol. 182:5127-5138(2000).
DR EMBL; AF135182; AAC09642.1; -
KW Plasmid.
SQ SEQUENCE 2376 AA; 262649 MW; B418DACE22DBH7BF CRC64;

Alignment Scores:
Pred. No.: 3e-28 Length: 2376
Score: 293.00 Matches: 65
Percent Similarity: 80.21% Conservatives: 12
Best Local Similarity: 67.71% Mismatches: 19
Query Match: 57.79% Indels: 1
DB: 2 Gaps: 0

US-09-856-221-3 (1-291) x Q9F9Z3 (1-2376)
QY 1 AATACCTTGCCTCAACATTACTGACGCGAGGATGACAGAGCAGCTGGCAGATTGCTGCAA 60
|||||
Db 1907 AnThrLeuLeuGlyIleThrGluArgGlnAspAlaGluAlaLeuAlaLysLeuLeuGln 1926
|||||
QY 61 ACTCAAGCAGTGAATAGCTTTGTCAGAGTATTAAATGACAGGC-AAGATGATTGCTGAA 119
|||||

Alignment Scores:
Pred. No.: 5.74e-19 Length: 2516
Score: 220.00 Matches: 50
Percent Similarity: 74.19% Conservatives: 19
Best Local Similarity: 53.76% Mismatches: 24
Query Match: 43.39% Indels: 1
DB: 2 Gaps: 0

US-09-856-221-3 (1-291) x Q9RN43 (1-2516)
QY 1 AATACCTTGCCTCAACATTACTGACGCGAGGATGACAGAGCAGCTGGCAGATTGCTGCAA 60
|||||
Db 2038 SerThrLeuGlnAsnIleIleGluArgGlnAspAlaGluAlaLeuAsnAlaLeuGln 2057
|||||
QY 61 ACTCAAGCAGTGAATAGCTTTGTCAGAGTATTAAATGACAGGC-AAGATGATTGCTGAA 119
|||||
Db 2058 AsnGlnAlaGluLeuIleLeuThrAsnLeuSerIleGlnAspLysThrIleGluGlu 2077
|||||
QY 120 ATTGATGCTGATGAAGTGGCGCTTAAGGAAGCCGTCATGTCACAAATCTGTTTGAC 179
|||||
Db 2078 LeuAspAlaGluLysThrValLeuGluLysSerLysAlaGlyAlaGlnSerArgPheAsp 2097
|||||
QY 180 AGCTTCAGTACGCTGTATGACGAAGATGTTAACTCCGCTGAAACAAACGATGATCTG 239
|||||
```



```
Ds 2098 SerTyrGlyLysLeuTyrAspGluAsnIleAsnAlaGlyGluAsnGlnAlaMetThrLeu 2117
QY 240 TATCTCTTCATCGTATGAGCAGCAGTACGGCC 278
Dd 2118 ArgAlaSerAlaAlaGlyLeuThrThrAlaValGlnAla 2130
RESULT 4
ID O85160 PRELIMINARY; PRT; 2504 AA.
AC O85160;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Insecticidal toxin complex protein TcBa.
GN TCBA.
OS Photobacterium luminescens (Xenorhabdus luminescens).
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Photobacterium.
OX NCBI_TaxID=29488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=W-14;
RX MEDLINE=98306208; PubMed=9641921;
RA Bowen D., Rocheleau T.A., Blackburn M., Andreev O., Golubeva E.,
RT Bhartiya R., French-Constant R.H.;
RT "Insecticidal toxins from the bacterium Photobacterium luminescens.";
RL Science 280:2129-2132(1998).
[2]
DR EMBL; AF047457; AAC38627.1;
DR EMBL; AF346498; AAL18460.1;
DR InterPro; IPR000566; Lipocin_CyTfABP.
DR PROSITE; PS00213; LIPOCALIN; UNKNOWN.1.
SQ SEQUENCE 2504 AA; 280654 MW; 8F29B1693D10470E CRC64;

Alignment Scores:
Pred. No.: 1,92e-17 Length: 2504
Score: 208.00 Matches: 46
Percent Similarity: 72.92% Conservative: 24
Best Local Similarity: 47.92% Mismatches: 26
Query Match: 41.03% Indels: 1
DB: 2 Gaps: 0

US-09-856-221-3 (1-291) x O85160 (1-2504)
QY 1 AATACCTTGCTCAACATTACTGACGCGCAGGATGACAGACCTGCGAGATTGCTGCA 60
Dd 2032 SerSerLeuLeuGlyTyrSerGluArgGlnAspAlaGluAlaMetSerGlnLeuLeuGln 2051
QY 61 ACTCAAGCGAGTGAATTACCTTTCAGAGTATTAAATGAGCGC-AAGATGATTCTGAA 119
Dd 2032 ThrGlnAlaSerGluLeuIleLeuThrSerIleArgMetGlnAspAsnGlnLeuAlaGlu 2071
QY 120 ATTGATGCTGATGAAGTGGCGGTTAAGGAAAGCCGTCATGTCACAAATCTGTTTGC 179
Dd 2072 LeuAspSerGluYsThrAlaLeuGlnValSerLeuAlaGlyValGlnGlnArgPheAsp 2091
QY 180 ACCTTCAGTACGCTGATGACGAGATGTTAATCCCGTGAAACACAGCGATGATCTG 239
Dd 2092 SerTyrSerGlnLeuTyrGluGluAsnIleAsnAlaGlyGluGlnArgAlaLeuAla 2111
QY 240 TATCTCTTCATCGTATGAGCAGCAGTACGGCCCTGCTGATG 287
Dd 2112 ArgSerGluSerAlaIleGluSerGlnGlyAlaGlnIleSerArgMet 2127
RESULT 5
Q93RN7
```

```

RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-W14; PubMed=11286884;
RX MEDLINE=21192684; PubMed=11286884;
RA Waterfield N.R., Bowen D.J., Fetherston J.D., Perry R.D.,
RA French-Constant R.H.;
RT "The tc genes of Photobacterium: a growing family.";
RL Trends Microbiol. 9:185-191(2001).
DR EMBL; AF047028; AAC38629.1; -;
DR EMBL; AF346499; AAU18472.1; -;
SQ SEQUENCE 1565 AA; 175717 MW; 66C3AB96C5FA0397 CRC64;

Alignment Scores:
Pred. No.: 7,04e-05 Length: 1565
Score: 109.00 Matches: 31
Percent Similarity: 56.52% Conservative: 21
Best Local Similarity: 33.70% Mismatches: 40
Query Match: 21.50% Indels: 1
DB: 2 Gaps: 0

US-09-856-221-3 (1-291) x 085156 (1-1565)
QY 7 TTGCTCAACATTACTTAAGCGCAGGATGCGAAGCACTGCGAGAAATGCTGCAAACTCAA 66
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1076 LeuLeuSerLeuLeuGluLeuArgSerGluArgAlaCysGlnGluGluLeuAlaGlnGln 1095
QY 67 GCGAGTGAATTAGCTTTGCGAGATATTAATGACAGGC-AAGATGATGCTGGAATTTGAT 125
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1096 LeuLeuAspMetSerSerTyrAlaIleThrLeuGlnGlnGlnAlaLeuAspGlyLeuAla 1115
QY 126 GCTGATGAAGTGGCGCTTAAGGAAGCGCTCATGCTGTCACCAATCTGTTTGGACAGCTTC 185
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1116 AlaAspArgLeuAlaLeuLeuAlaSerGlnAlaThrAlaGlnGlnArgHisAspHisTyr 1135
QY 186 AGTAGCTGTATGACGAAGATTTACTCGGTGAAACAAAGCAGGATGATGATATCTC 245
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1136 TyrThrLeuTyrGlnAsnAsnIleSerSerAlaGluGlnLeuValMetAspThrGlnThr 1155
QY 246 TCTTCATCGGTATGACGACGACGATGACGCGCTG 281
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1156 SerAlaGlnSerLeuIleSerSerThrGlyVal 1167

RESULT 7
Q99Y37 PRELIMINARY; PRT; 296 AA.
AC Q99Y37;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Similar to several eukaryotic hypersensitive-induced response
DE proteins.
GN Spi1884.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SF370 / ATCC 700294 / SEROTYPE M1;
RX MEDLINE=21192684; PubMed=11296296;
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
DR EMBL; AE006613; AAK34596.1; -;
DR InterPro; IPR001107; Band_7.
DR Pfam; PF01145; Band_7; 1.
DR SMART; SM00244; PHB; 1.
KW Complete proteome.
SQ SEQUENCE 296 AA; 33066 MW; 530DA0BFB73E7340 CRC64;

```

```

Alignment Scores:
Pred. No.: 4.02 Length: 296
Score: 71.00 Matches: 25
Percent Similarity: 46.67% Conservative: 17
Best Local Similarity: 27.78% Mismatches: 38
Query Match: 14.00% Indels: 10
DB: 16 Gaps: 2

US-09-856-221-3 (1-291) x Q99Y37 (1-296)
QY 36 AGAGCACTGGCAGAAATGCTGCAAACTCAAGGCAGTGAATTAGCTTGCAGAGATATTA 95
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 175 GlnSerMetAsnGluIleAsnAlaAlaGlnArgLysArgValAlaAlaGlnGluLeuAla 194
QY 96 AATGACGACGAAGATGATGCTGAAATGATGCTGAAGTGGCGCTTAAGGAAAGCCGT 155
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 195 AsnAlaAspLysIleLysIleValThrAlaAlaGluAlaGluLysAspArgLeu 214
QY 156 CATGGT-----GCACAACTCGT-----TTTGACAGCTTC 185
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 215 HisGlyValGlyIleAlaGlnArgLysAlaIleValAspGlyLeuAlaGluSerIle 234
QY 186 AGTAGCTGTATGACGAAGATGTTAACTCCGTCGTAACAAAGCAGGATGATGATCTC 245
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 235 GlnGluLeuLysGluAlaAsnIleSerLeuAsnGluGlnIleMetSerIleLeuLeu 254
QY 246 TCTTCATCGGTATGACGACGACGATGACG 275
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 255 ThrAsnGlnTyrLeuAspThrLeuAsnThr 264

RESULT 8
Q9F5F7 PRELIMINARY; PRT; 1693 AA.
AC Q9F5F7;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Riorf93 protein.
GN RIORF93.
OS Agrobacterium rhizogenes.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=359;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MAFF03-01724;
RA Moriguchi K., Maeda Y., Satou M., Satuti N., Kataoka M., Tanaka N.,
RA Yoshida K.;
RT "The complete nucleotide sequence of a R1 (root inducing) plasmid
RT indicates its chimerical structure between Ti and Sym plasmids.";
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-MAFF03-01724;
RA Moriguchi K., Maeda Y., Satou M., Kataoka M., Tanaka N., Yoshida K.;
RT "Analysis of unique variable region of a plant root inducing plasmid,
RT pRi1724, by the construction of its physical map and library.";
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-MAFF03-01724;
RA Moriguchi K., Nishida T., Maeda Y., Tanaka N., Yoshida K.;
RT "Genome structure of R1 plasmid (1): Construction of linking library
RT and physical map of pRi1724 in Japanese Agrobacterium.";
RL Nucleic Acids Symp. Ser. 39:189-190(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-MAFF03-01724;
RX MEDLINE=20241294; PubMed=10780382;
RA Maeda Y., Moriguchi K., Kataoka M., Satou M., Satuti N., Tanaka N.,
RA Yoshida K.;
RT "Genome structure of R1 plasmid (1): Sequencing analysis of T-DNA and

```

RT its flanking regions of pRI1724 in Japanese Agrobacterium  
 RL rhizogenes<sup>+</sup>;  
 RL Nucleic Acids Symp. Ser. 42:67-68(1999).  
 DR EMBL; AP002086; BAB16212.1; -;  
 DR HSSP; P14385; 2ADM.

DR InterPro; IPR001410; DEAD.  
 DR InterPro; IPR001650; Helicase\_C.  
 DR InterPro; IPR002296; N12N6\_mtfase.  
 DR InterPro; IPR002052; N6\_Mtase.  
 DR InterPro; IPR000051; N6\_Mtase.  
 DR Pfam; PF00271; helicase\_C; 1.  
 DR PRINTS; PR00507; N12N6MTFRASE.  
 DR SMART; SM00487; DEXDC; 1.  
 DR PROSITE; PS00092; N6\_MTASE; UNKNOWN\_1.  
 KW Plasmid.

SQ SEQUENCE 1693 AA; 188083 MW; 62B8E7E786782555 CRC64;

Alignment Scores:  
 Pred. No.: 4.8 Length: 1693  
 Score: 71.00 Matches: 21  
 Percent Similarity: 54.10% Conservative: 12  
 Best Local Similarity: 34.43% Mismatches: 26  
 Query Match: 14.00% Indels: 2  
 DB: 2 Gaps: 2

US-09-856-221-3 (1-291) x Q9F5F7 (1-1693)

QY 111 ATTCTGAATTCATGCTGATCAAGTCGGCTTAAGCAAGCCGTCATGTCACAATCT 170  
 DB 1033 lilealgluileglyvalaspinllelevalaspglualaglnphearglysleu 1052  
 QY 171 CGTTTT--GACAGCTTCAGCTACGCTGATGACGAAGATGTTAACTCCGGTGAACAA 227  
 DB 1053 SerPheAlaThrAsnMetSerThrLeuLysGlyIleAspProAsn--GlySerGlnArg 1071  
 QY 228 CGGATGATCTGATCTCTCTTCATGCTGATGACGACGACGATGACGCCCTGCTATG 287  
 DB 1072 AlaTrpAspLeuIrrValLysSerArgPheLeuGluThrLysAsnProGlyArgAlaLeu 1091  
 QY 288 GTG 290  
 DB 1092 Val 1092

RESULT 9  
 Q07893  
 ID Q07893 PRELIMINARY; PRT: 1379 AA.  
 AC Q07893  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE Protoporphytin IX:Mg chelatease.  
 GN OLIVE.  
 OS Antirrhinum majus (Garden snapdragon).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Asteridae; euasterids I; Lamiales; Veronicaceae; Antirrhinum.  
 OX NCBI\_TaxID=4151;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=WILD TYPE, AND CV. J.I.98;  
 RX MEDLINE=94008977; PubMed=8404842;  
 RA Hudson A., Carpenter R., Doyle S., Coen E.S.;  
 RT "Olive: a key gene required for chlorophyll biosynthesis in  
 Antirrhinum majus";  
 RL EMBO J. 12:3711-3719(1993).  
 DR EMBL; X73144; CAA51664.1; -;  
 DR InterPro; IPR003672; COB/Mg-chlase.  
 DR Pfam; PF02514; COB-Mg\_chel; 1.  
 SQ SEQUENCE 1379 AA; 153217 MW; 0088E819984F505D CRC64;

Alignment Scores:  
 Pred. No.: 8.45 Length: 1379  
 Score: 69.00 Matches: 24

Percent Similarity: 40.96% Conservative: 10  
 Best Local Similarity: 28.92% Mismatches: 25  
 Query Match: 13.61% Indels: 24  
 DB: 10 Gaps: 3  
 US-09-856-221-3 (1-291) x Q07893 (1-1379)

QY 99 GCAGGCAAGATGCTGCTGAATGAT----- 125  
 DB 1102 AlaValLysMetValAlaGluLeuAspGluProValGluGlnAsnPheValArgLysHis 1121  
 QY 126 -----GCTGATGAAGTGGCGCTTAAGGAAAGCGTCATGTCACAATCTCGT 173  
 DB 1122 AlaLeuGluGlnAlaLysGluLeuGlyVal---GluValArgGluAlaAlaSerArgIle 1140  
 QY 174 TTTCAGAGCTTCAGTACGCTGATGACGAAGATGTTAAC----- 212  
 DB 1141 PheSerAsnAlaSerGlySerTyrSerAsnIleAsnLeuAlaValGluAsnSerSer 1160  
 QY 213 ---TCCGGTCAAAAACAGCGATGATGATCTCTTCATCGGTATGAGCACACG 269  
 DB 1161 TrpAsnAspGluLysGlnLeuGlnAspMetTyrLeuSerArgLysSerPheAlaPheAsp 1180  
 QY 270 AGTACGGCC 278  
 DB 1181 SerAspAla 1183

RESULT 10  
 Q16734  
 ID Q16734 PRELIMINARY; PRT: 299 AA.  
 AC Q16734;  
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE K09F6.7 protein.  
 GN K09F6.7.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Pelodierinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BRISTOL N2;  
 RX MEDLINE=94150718; PubMed=7906398;  
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M., Coulson A.,  
 RA Craxton M., Dear S., Du Z., Durbin R., Favell A., Fulton L.,  
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Fuiton L.,  
 RA Jones A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,  
 RA Gardiner M., Kershaw J., Kirschen J., Laister N., Latreille P.,  
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,  
 RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,  
 RA Smailson N., Smith A., Sonnenhammer E., Staden R., Sulston J.,  
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,  
 RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;  
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 elegans";  
 RN Nature 368:32-38(1994).  
 RP [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BRISTOL N2;  
 RX Sheet P.;  
 RT "The sequence of C. elegans cosmid K09F6.";  
 RL submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BRISTOL N2;  
 RA Waterston R.;  
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.  
 DR EMBL; AT016683; AAB66197.1; -;  
 DR HSSP; P15919; IRMD.  
 DR InterPro; IPR001841; Znf\_ring.  
 DR Pfam; PF00097; zf-C3HC4; 1.



```

DB:          3          Gaps:          2
US-09-856-221-3 (1-291) x 059966 (1-626)
QY 60 AACTCAGGCGAGTGAATAGCTTTCCAGATTTAAATACGAGCAAGATGATT----- 113
   |||||
Db 344 AsnSerArgGlnArgAlaSerGluAlaAspArgAlaHisIleGlyHisValIleProArg 363
QY 114 --GCTGAAATGATCTCATCAAGTCGCTTAAGGAAAGCCGTCATGGTCACAATCT 170
   |||||
Db 364 TrpLeuGluIleAlaGluTrpIleAlaLeuAspGluSerGln-----GlnHis 380
QY 171 CGTTTTCAGAGCTTCAGTCAGCTGTGTGACGAGAGATGTTAACTCCGCTGAAACAAACGCG 230
   |||||
Db 381 GlnAspValAsnHisSerGluLeuTrpSerValTrpLeuAsnArgMetAspIleGlnThr 400
QY 231 ATGGATCTGATCTCTCTTCATCGCTATTG 260
   |||||
Db 401 TyrAspIleHisTyrAlaGlyPheAlaLeu 410

RESULT 13
ID O65808 PRELIMINARY; PRT; 1383 AA.
AC O65808;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Magnesium chelatase subunit.
GN CHLH.
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
OX NCBI_TaxID=3847;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RESNIK;
RA Nakayama M., Masuda T., Bando T., Ohta H., Takamiya K.I.;
RT "Cloning and characterization of the soybean chlH gene coding for a
RL subunit of magnesium-chelatase.";
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ001091; CAA04526.1;
DR InterPro; IPR003672; COB/Mg-chltase.
DR Pfam; PF02514; COB/Mg-chel. 1.
SQ SEQUENCE 1383 AA; 153492 MW; 26C9F46515205E75 CRC64;

Alignment Scores:
Pred. No.:      11.3      Length:      1383
Score:          68.00     Matches:      24
Percent Similarity: 42.68%   Conservative: 11
Best Local Similarity: 29.27% Mismatches:    25
Query Match:      13.41%   Indels:       22
DB:             10      Gaps:         3

US-09-856-221-3 (1-291) x O65808 (1-1383)
QY 99 GCAGCGAAGATGATTCTGAAATTCAT-----GCTGATCAAGTCGGCTTAAG----- 146
   |||
Db 1106 AlaValIleMetValAlaGluLeuAspGluProAlaGluGlnAsnTyrValIleTyrHis 1125
QY 147 -----GAAAGCCGTCATGTCGACCAATCTCGTTT 176
   |||
Db 1126 AlaSerGluGlnAlaGlnAlaLeuGlyValGluValArgGluAlaAlaThrArgIlePhe 1145
QY 177 GACAGTTCATGCTAGCTGTATGACGAGATGTTAAC----- 212
   |||
Db 1146 SerAsnAlaSerGlySerTyrSerAsnIleAsnLeuAlaValGluAsnSerSerTrp 1165
QY 213 TCCGCTGAAACAAACGATGATGATCTCTTCATCGGTATTGACGACACGACGT 272
   |||
Db 1166 AsnAspGluIleGlnLeuGlnAspMetTyrLeuSerArgIleGlySerPheAlaPheAspSer 1185
QY 273 ACGGCC 278

```

```

|||
Db 1186 AspAla 1187
RESULT 14
QYR612 PRELIMINARY; PRT; 1693 AA.
AC QYR612;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE TIOF47 protein.
GN TIOF47.
OS Agrobacterium tumefaciens.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=358;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF301001;
RX MEDLINE=20184752; PubMed=10721727;
RA Suzuki K., Hattori Y., Uraji M., Iwata K., Murata K.,
RA Katoh A., Yoshida K.;
RT "Complete nucleotide sequence of a plant tumor-inducing Ti plasmid.";
RL Gene 242:331-336(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF301001;
RX MEDLINE=98193120; PubMed=9524202;
RA Suzuki K., Ohta N., Hattori Y., Uraji M., Katoh A., Yoshida K.;
RT "Novel structural difference between nopaline- and octopine- type trbJ
RT gene: construction of genetic and physical map and sequencing of
RT trbJ and rep gene clusters of a new Ti plasmid pTi-SAKURA.";
RL Blochim. Biophys. Acta 1396:1-7(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF301001;
RX MEDLINE=98193120; PubMed=9524202;
RA Hattori Y., Suzuki K., Ohta N., Uraji M., Katoh A., Yoshida K.;
RT "Genome structure of pTi-SAKURA (Ti): Strategy for DNA sequencing of a
RT Japanese cherry-Ti plasmid.";
RL Nucleic Acids Symp. Ser. 37:159-160(1998).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF301001;
RX Ohta N., Suzuki K., Hattori Y., Uraji M., Katoh A., Yoshida K.;
RT "Genome structure of pTi-SAKURA (iii): Characteristics of T-DNA.";
RL Nucleic Acids Symp. Ser. 39:185-186(1998).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF301001;
RX Uraji M., Suzuki K., Ohta N., Hattori Y., Katoh A., Yoshida K.;
RT "Genome structure of pTi-SAKURA (IV): Characteristics of tra region.";
RL Nucleic Acids Symp. Ser. 39:187-188(1998).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF301001;
RX Hattori Y., Suzuki K., Ohta N., Uraji M., Katoh A., Yoshida K.;
RT "Genome structure of pTi-SAKURA (V): Complete nucleotide sequence of
RT plasmid pTi-SAKURA's vir region in Agrobacterium tumefaciens.";
RL Nucleic Acids Symp. Ser. 39:265-266(1998).
DR EMBL; AB016260; BAA87672.1;
DR HSP; P14385; 2ADM.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR002296; N12N6_mtfase.
DR InterPro; IPR002052; N6_Mtase.
DR Pfam; PF002051; SAM_bind.
DR PRINTS; PR00507; N12N6MTFRASE.
DR PROSITE; PS00092; N6_MTASE; UNKNOWN_1.
KW ATP-binding; Helicase; plasmid.
SQ SEQUENCE 1693 AA; 187978 MW; 0DD679AFBE6514BE CRC64;

```

Alignment Scores:

[illegible]

RESULT	ID	Q8U634	PRELIMINARY:	PRT:	1693	AA.
AC	Q8U634:					
DT	01-JUN-2002	(TrEMBLrel. 2L, Created)				
DT	01-JUN-2002	(TrEMBLrel. 2L, Last sequence update)				
DT	01-JUN-2002	(TrEMBLrel. 2L, Last annotation update)				
DE	Helicase, SNF2 family.					
DE	AGU6101 OR AGR_PTI_188.					
GN	Agrobacterium tumefaciens (strain C58 / ATCC 33970).					
OG	Plasmid pTiC58.					
OC	Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;					
OC	Rhizobiaceae; Rhizobium.					
ON	NCBI_TaxID=176299;					
OX	[1]					
RP	SEQUENCE FROM N.A.					
RP	MEDLINE=21608550; PubMed=11743193;					
RA	Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,					
RA	Kura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,					
RA	Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,					
RA	Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,					
RA	Kutyavin T., Levy R., Li M.-J., McLelland E., Palmieri A., Gordon D.,					
RA	Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,					
RA	Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,					
RA	Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,					
RA	Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,					
RA	Nester E.W.;					
RT	"The genome of the natural genetic engineer Agrobacterium tumefaciens					
RT	C58.";					
RT	Science 294:2317-2323(2001).					
RN	[2]					
RN	SEQUENCE FROM N.A.					
RP	MEDLINE=21608551; PubMed=11743194;					
EX	Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,					
RA	Qurollio B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,					
RA	Houmello K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,					
RA	Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,					
RA	Flanagan C., Crowell C., Gursion J., Lomo C., Sear C., Strub G.,					
RA	Cielo C., Slater S.;					
RT	"Genome sequence of the plant pathogen and biotechnology agent					
RT	Agrobacterium tumefaciens C58.";					
RT	Science 294:2323-2328(2001).					
RL	EMBL; AE009429; AAL46337.1;					
DR	EMBL; AE007937; AAK91062.1;					
DR	EMBL; Complete proteome.					
KW	Plasmid; Complete proteome.					
SC	SEQUENCE 1693 AA; 188061 MW; B5F2E130AA696DFB CRC64;					
Alignment Scores:						
pred. No.:		11.6				Length: 1693
Score:		68.00				Matches: 19
Percent Similarity:		56.60%				Conservative: 21
Best Local Similarity:		59.95%				Mismatches: 11

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model  
Run on: January 15, 2003, 15:39:33 ; Search time 30.7357 Seconds  
(without alignments)  
2878.693 Million cell updates/sec

Title: US-09-856-221-4  
Perfect score: 605  
Sequence: 1 gtagagcgagcattgtcaa.....atttggaagagatgaacgg 332

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 1816940

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlpl  
-O=/cgn2.1/USPTO.spool/US09856221/runat\_15012003\_153924\_12438/app\_query.fasta\_1.1948  
-DB=A\_Geneseq\_101002 -QFMT=fastan -SUFFIX=rag -MINMATCH=0.1 -LOOPCL=0  
-LIST=45 -DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09856221.ecgn\_1.1.60 -runat\_15012003\_153924\_12438 -NCPU=6 -ICPU=3  
-NO\_XLPHY -NO\_MMAPP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

A\_Geneseq\_101002.\*  
1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	446	73.7	579	18	AAW18301
2	446	73.7	579	19	AAW56574
3	446	73.7	2516	18	AAW17899
4	446	73.7	2516	19	AAW56572
5	446	73.7	2516	22	AAW72609
6	446	73.7	2517	22	AAW72611
7	446	73.7	2537	22	AAW72614
8	440	72.7	2522	20	AAW33729
9	418	69.1	573	18	AAW18303
10	418	69.1	573	19	AAW56559
11	418	69.1	2504	22	AAW56557
12	418	69.1	2504	22	AAW72610
13	418	69.1	2505	22	AAW72612
14	410	67.8	2504	18	AAW17871
15	395.5	65.4	2376	22	AAW97694
16	268	44.3	1565	18	AAW18305
17	268	44.3	1565	19	AAW56568
18	174.5	28.8	562	18	AAW17886
19	174.5	28.8	562	19	AAW56545
20	174.5	28.8	1189	18	AAW17884
21	174.5	28.8	1189	19	AAW56543
22	79	13.1	250	22	ABW67792
23	73.5	12.1	173	22	AAW25690
24	70.5	11.7	507	17	AAW00290
25	70.5	11.7	532	17	AAW00291
26	69.5	11.5	467	21	AAW01601
27	69.5	11.5	467	21	AAW75572
28	69.5	11.5	467	22	AAW68934
29	69.5	11.5	467	22	AAW68935
30	69.5	11.5	601	23	ABW28734
31	69.5	11.5	687	19	AAW41586
32	69.5	11.5	1392	20	AAW06999
33	69.5	11.5	1427	12	AAW10534
34	69	11.4	254	21	AAW29613
35	68	11.6	582	22	AAU03896
36	68	11.2	696	22	AAU33947
37	68	11.2	698	22	AAU36732
38	68	11.2	1056	23	ABW35622
39	67.5	11.2	330	22	AAE01417
40	67.5	11.2	416	18	AAW55613
41	67.5	11.2	635	18	AAW55723
42	67.5	11.2	646	18	AAW55314
43	67.5	11.2	668	18	AAW55709
44	67.5	11.2	680	19	AAW46516
45	67.5	11.2	704	20	AAW17210

ALIGNMENTS

RESULT 1  
AAW18301  
ID AAW18301 standard; Protein; 579 AA.  
XX  
AC AAW18301;  
XX  
DT 29-JAN-1998 (first entry)  
XX  
DE Photorhabdus luminescens insect toxin TcdA11.  
XX  
KW Insecticide; insect; pest control; biological control;  
KW Photorhabdus luminescens; TcdA; Southern corn rootworm;  
KW Colorado potato beetle; Western corn rootworm; meal worm;  
KW boll weevil; turf grub; Coleoptera; beet armyworm; black cutworm;  
KW cabbage looper; codling moth; corn earworm; European corn borer;  
KW tobacco hornworm; tobacco budworm; Lepidoptera; Hymenoptera;  
KW Diptera; Dictyoptera; Acarina; Homoptera.  
XX  
OS Photorhabdus luminescens strain W-14 (ATCC 55397).

XX W09717432-A1.  
PN 15-MAY-1997.  
PD 06-NOV-1996; 96WO-US18003.  
PE 28-AUG-1996; 96US-0705484.  
PR 06-NOV-1995; 95US-0007255.  
PR 28-FEB-1996; 96US-0608423.  
XX (WISC ) WISCONSIN ALUMNI RES FOUND.  
XX Blackburn MB, Bowen DJ, Cliche TA, Ensign JC, Fatig R;  
PI French-Constant RH, Guo L, Hey TD, Merlo DJ, Orr GL;  
PI Petell J, Roberts JL, Rocheleau TA, Schoonover S;  
PI Strickland JA;  
XX WPI; 1997-281022/25.  
DR N-PSDB; AAT68845.  
XX Photorhabdus sp. insecticidal protein toxins and DNA encoding them -  
PT can be genetically engineered into insect larvae food and plants for  
PT insect control  
XX Claim 34; Page 213-215; 276pp; English.  
XX This polypeptide comprises the 63.6 kDa TcdAIII insecticidal toxin  
CC protein of Photorhabdus luminescens W-14. Its sequence was deduced  
CC from a genomic DNA clone (AAT68845) and includes N-terminal and  
CC tryptic peptide sequences obtained from the isolated protein.  
CC TcdAIII is a proteolytic cleavage product of tcdA (see AAW17899).  
CC Claimed toxins of P. luminescens (see AAW17871, AAW17884-89,  
CC AAW17899-900, AAW18301-06) can be produced by recombinant DNA methods and  
CC applied to, or genetically engineered into, insect larvae food and  
CC plants for insect control. The toxins are particularly effective  
CC against Southern corn rootworm, Colorado potato beetle, Western corn  
CC rootworm, meal worm, boll weevil and turf grub (Coleoptera), beet  
CC armyworm, black cutworm, cabbage looper, codling moth, corn earworm,  
CC European corn borer, tobacco hornworm and tobacco budworm  
CC (Lepidoptera), and are also active against insects of the orders  
CC Hymenoptera, Diptera, Dictyoptera, Acarina and Homoptera. (All  
CC claimed).  
XX Sequence 579 AA;  
SQ  
Alignment Scores:  
Pred. No.: 1.77e-50 Length: 579  
Score: 446.00 Matches: 86  
Percent Similarity: 89.09% Conservative: 12  
Best Local Similarity: 78.18% Mismatches: 12  
Query Match: 73.72% Indels: 0  
DB: 18 Gaps: 0  
US-09-856-221-4 (1-332) x AAW18301 (1-579)  
QY 3 GAAGGGCAGTATTCGAAACAACTATCTGGAAACCCACAGGACAACTCAGGCACAG 62  
DB 281 GluAlaIaValLeuGlnLysThrSerLeuLysThrGlnGlnGlnThrGlnSerGln 300  
QY 63 CTGGGCTTCCTACAAAGCAATTCAGCAATACAGGCTGTGTATTAAGTGGCTGGCGGA 122  
DB 301 LeuAlaPheLeuGlnArgLysPheSerAsnGlnAlaLeuTyrAsnTTPLeuArgGlyArg 320  
QY 123 TTGGGGCTATTATATATCAGTTTATGACTTGGCTGTGTTCCTGTGTGTTCATGGCTGAA 182  
DB 321 LeuAlaIaIleTyrPheGlnPheTyrAspLeuAlaIaIaArgCysLeuMetAlaGlu 340  
QY 183 CAAACTTACAGTATGATTAAGACGATTAAGCTGTACGCTTCATTAAAGCCCGGTGGCTGG 242  
DB 341 GlnAlaTyrArgTTPGluLeuAsnAspSerAlaArgPheIleLysProGlyAlaTTP 360  
QY 243 CATGGCACTATGCTGTTTGTAGCAGGTGAACCTTCATGCTGCTGAATTTGGCACAGTC 302

DB 361 GlnGlyThrTyrAlaGlyLeuLeuAlaGlyGluThrLeuMetLeuSerLeuAlaGlnMet 380  
QY 303 GAAAAAACTATTGGAAAAAGATGAACGG 332  
DB 381 GluAspAlaHisLeuLysArgAspLysArg 390  
RESULT 2  
AAW56574  
ID AAW56574 standard; Protein; 579 AA.  
XX  
AC AAW56574;  
XX  
DT 07-AUG-1998 (first entry)  
XX Toxin TcdAIII, encoded by the tcdA gene from genomic region tcd.  
DE Photorhabdus luminescens W-14; nematode; symbiotic;  
KW Heterorhabditis; tca; tcb; tcc; tcd; insecticidal activity; toxin;  
KW Lepidoptera; Coleoptera; Hymenoptera; Diptera; Dictyoptera; Acarina;  
KW Homoptera; Southern; Western corn rootworm; Colorado potato beetle;  
KW mealworm; boll weevil; turf grub; beetle armyworm; black cutworm;  
KW cabbage looper; codling moth; corn earworm; European corn borer;  
KW Tobacco hornworm; budworm.  
XX Photorhabdus luminescens.  
OS  
XX W09808932-A1.  
PN  
XX 05-MAR-1998.  
PD  
XX 05-MAY-1997; 97WO-US07657.  
PF  
XX 06-NOV-1996; 96WO-US18003.  
PR 28-AUG-1996; 96US-0705484.  
PR 06-NOV-1996; 96US-0743699.  
XX (DOWC ) DOWELANCO.  
PA (WISC ) WISCONSIN ALUMNI RES FOUND.  
PA Blackburn MB, Bowen DJ, Cliche TA, Ensign JC, Fatig R;  
PI French-Constant RH, Guo L, Hey TD, Merlo DJ, Orr GL;  
PI Petell J, Roberts JL, Rocheleau TA, Schoonover S;  
PI Strickland JA, Sukhapingda K;  
XX WPI; 1998-179427/16.  
DR N-PSDB; AAV29931.  
XX Isolated toxins from Photorhabdus luminescens strains - useful for  
PT control of insect pests  
XX Claim 34; Pages 245-247; 321pp; English.  
XX The present sequence represents a protein named TcdAIII of the bacterium  
CC Photorhabdus luminescens (W-14). This is a symbiotic bacterium of the  
CC nematodes of the Heterorhabditis genus. The bacterium has at least 4  
CC distinct genomic regions, tca, tcb, tcc, and tcd. Peptide products are  
CC produced from these regions that are associated with insecticidal  
CC activity. The native toxins are secreted proteins. The proteins are  
CC toxic to insects upon exposure and especially when ingested. The  
CC nucleic acid sequence can be used to produce transgenic plants,  
CC baculoviruses or microbial hosts for toxin production. They can be used  
CC to control insects pests from the Lepidoptera, Coleoptera, Hymenoptera,  
CC Diptera, Dictyoptera, Acarina or Homoptera orders, especially the  
CC Southern or Western corn rootworm, Colorado potato beetle, mealworm,  
CC boll weevil, turf grub, beetle armyworm, black cutworm, cabbage looper,  
CC codling moth, corn earworm, European corn borer or tobacco hornworm  
CC or budworm.  
XX Sequence 579 AA;  
SQ  
Alignment Scores:  
Pred. No.: 1.77e-50 Length: 579



```
Score: 446.00 Matches: 86
Percent Similarity: 89.09% Conservative: 12
Best Local Similarity: 78.18% Mismatches: 12
Query Match: 73.72% Indels: 0
DB: 19 Gaps: 0

US-09-856-221-4 (1-332) x AAW56574 (1-579)

QY 3 GAAGCGCAGTATTGCAAAACAACTATCTGGAACCCACAGGCACAACTCAGGCACAG 62
DB 281 GluAlaValLeuGlnLysThrSerLeuLysThrGlnGlnGluGlnThrGlnSerGln 300
QY 63 CTGGCCTTCTACAAAGCAAAATTCAGCAATACAGCGTTGTATACATGGCTAGCTGGCGCA 122
DB 301 LeuAlaPheLeuGlnArgLysPheSerAsnGlnAlaLeuLysThrLeuArgGlyArg 320
QY 123 TTGGCGGCTATTATTATCAGTTTATGACTTGGCTGTTCCCTGTTGTTGATGCTGAA 182
DB 321 LeuAlaAlaIleTyrPheGlnPheTyrAspLeuAlaValAlaIleArgCysLeuMetAlaGlu 340
QY 183 CAATCTACAGATGAAATGACGATAAAGCTGTACGCTTCATTAAAGCCCGTGCCTGG 242
DB 341 GlnAlaTyrArgTrpGluLeuAsnAspSerAlaArgPheIleLysProGlyAlaTrp 360
QY 243 CATGGCACTATGCTGTTGTTAGCAGTGAACCTTGTGATGCTGAATTTGGCAGATG 302
DB 361 GlnGlyThrTyrAlaGlyLeuLeuAlaGlyGluThrLeuMetLeuSerLeuAlaGlnMet 380
QY 303 GAAACAACTATTGGAAAGATGACGCG 332
DB 381 GluAspAlaHisLeuLysArgAspLysArg 390

RESULT 3
AAW17899
ID AAW17899 standard; Protein; 2516 AA.
XX
AC AAW17899;
XX
DT 29-JAN-1998 (first entry)
XX
DE Photobhabdus luminescens insect toxin TcdA.
XX
KW Insecticide; insect; toxin; pest control; biological control;
KW Photobhabdus luminescens; TcdA; Southern corn rootworm;
KW Colorado potato beetle; Western corn rootworm; meal worm;
KW boll weevil; turf grub; Coleoptera; beet armyworm; black cutworm;
KW cabbage looper; codling moth; corn earworm; European corn borer;
KW tobacco hornworm; tobacco budworm; Lepidoptera; Hymenoptera;
KW Diptera, Dictyoptera; Acarina; Homoptera.
XX
OS Photobhabdus luminescens strain W-14 (ATCC 55397).
XX
FH
FT Key Location/Qualifiers
FT Protein 1..2516
FT /label= TcdA
FT Protein 89..1937
FT /label= TcdAII
FT Peptide 89..100
FT /note= "S2 N-terminus (Claim 30)"
FT Peptide 284..299
FT /note= "tryptic peptide (Claim 30)"
FT Peptide 554..563
FT /note= "tryptic peptide (Claim 30)"
FT Peptide 1080..1092
FT /note= "isolated N-terminal peptide (Claim 30)"
FT Peptide 1385..1400
FT /note= "tryptic peptide (Claim 30)"
FT Peptide 1478..1497
FT /note= "tryptic peptide (Claim 30)"
FT Peptide 1620..1642
FT /note= "claimed peptide (Claim 30)"
FT Peptide 1938..1948
FT /note= "N-terminal peptide (Claim 30)"
FT

Alignment Scores:
Pred. No.: 2,75e-50 Length: 2516
Score: 446.00 Matches: 86
Percent Similarity: 89.09% Conservative: 12
Best Local Similarity: 78.18% Mismatches: 12
Query Match: 73.72% Indels: 0
DB: 18 Gaps: 0

US-09-856-221-4 (1-332) x AAW17899 (1-2516)

QY 3 GAAGCGCAGTATTGCAAAACAACTATCTGGAACCCACAGGCACAACTCAGGCACAG 62
DB 2218 GluAlaValLeuGlnLysThrSerLeuLysThrGlnGlnGluGlnThrGlnSerGln 2237
QY 63 CTGGCCTTCTACAAAGCAAAATTCAGCAATACAGCGTTGTATACATGGCTAGCTGGCGCA 122
DB 2238 LeuAlaPheLeuGlnArgLysPheSerAsnGlnAlaLeuLysThrLeuArgGlyArg 2257
QY 123 TTGGCGGCTATTATTATCAGTTTATGACTTGGCTGTTCCCTGTTGTTGATGCTGAA 182
DB 2258 LeuAlaAlaIleTyrPheGlnPheTyrAspLeuAlaValAlaArgCysLeuMetAlaGlu 2277
```

QY 183 CAACCTTACCATGATTAAGCTGTAAGCTTCAATTAAGCCGCTGCTGG 242  
 DB 2278 GlnAlaIyArGrpGluLeuAsnAspSerAlaArGpHeIIeIySProGlyAlaItrp 2297  
 QY 243 CATGCCATTATGCTGTTTGTAGCAGGTGAACCTTGTGCTGAATTTGGCAGATG 302  
 DB 2298 GlnGlyThrTyrAlaGlyLeuLeuAlaGlyGluThrLeuMetLeuSerLeuAlaGlnMet 2317  
 QY 303 GAAAAAACTATTTGGAAGAAGATGAACG 332  
 DB 2318 GluAspAlaHisLeuIySArgAspIySArg 2327

RESULT 4  
 AAW56572  
 ID AAW56572 standard; Protein: 2516 AA.  
 XX  
 AC AAW56572;  
 XX  
 DT 07-AUG-1998 (first entry)  
 DE Toxin TcdA, encoded by the tcdA gene from genomic region tcd.  
 XX  
 KW Photorhabdus luminescens W-14; nematode; symbiotic;  
 KW Heterorhabditis; tcd; tcd; insecticidal activity; toxin;  
 KW Lepidoptera; Coleoptera; Hymenoptera; Diptera; Dictyoptera; Acarina;  
 KW Homoptera; Southern; Western corn rootworm; Colorado potato beetle;  
 KW mealworm; boll weevil; turf grub; beetle armyworm; black cutworm;  
 KW cabbage looper; codling moth; corn earworm; European corn borer;  
 KW Tobacco hornworm; budworm.  
 XX  
 OS Photorhabdus luminescens.  
 XX  
 PN W09808932-A1.  
 XX  
 PD 05-MAR-1998.  
 XX  
 PF 05-MAY-1997; 97WO-US07657.  
 XX  
 PR 06-NOV-1996; 96WO-US18003.  
 PR 28-AUG-1996; 96US-0705484.  
 PR 06-NOV-1996; 96US-0743699.  
 XX  
 PA (DOWC ) DOWELANCO.  
 PA (WISC ) WISCONSIN ALUMNI RES FOUND.  
 XX  
 PI Blackburn MB, Bowen DJ, Cliche TA, Ensign JC, Fatig R;  
 PI French-Constant RH, Guo L, Hey TD, Merlo DJ, Orr GL;  
 PI Petell J, Roberts JL, Rocheleau TA, Schoonover S;  
 PI Strickland JA, Sukhapinda K;  
 XX  
 WPI: 1998-179427/16.  
 DR N-PSDB; AAW29928.  
 XX  
 PT Isolated toxins from Photorhabdus luminescens strains - useful for  
 PT control of insect pests  
 XX  
 PS Claim 34; Pages 224-231; 321pp; English.  
 XX  
 CC The present sequence represents a protein named TcdA of the bacterium  
 CC Photorhabdus luminescens (W-14). This is a symbiotic bacterium of the  
 CC nematodes of the Heterorhabditis genus. The bacterium has at least 4  
 CC distinct genomic regions, tcd, tcb, tcc, and tcd. Peptide products are  
 CC produced from these regions that are associated with insecticidal  
 CC activity. The native toxins are secreted proteins. The proteins are  
 CC toxic to insects upon exposure and especially when ingested. The  
 CC nucleic acid sequence can be used to produce transgenic plants,  
 CC baculoviruses or microbial hosts for toxin production. They can be used  
 CC to control insects pests from the Lepidoptera, Coleoptera, Hymenoptera,  
 CC Diptera, Dictyoptera, Acarina or Homoptera orders, especially the  
 CC Southern or Western corn rootworm, Colorado potato beetle, mealworm,  
 CC boll weevil, turf grub, beetle armyworm, black cutworm, cabbage looper,  
 CC codling moth, corn earworm, European corn borer or tobacco hornworm

CC or budworm.  
 XX Sequence 2516 AA;  
 SQ  
 Alignment Scores: 2.75e-50 Length: 2516  
 Pred. No.: 446.00 Matches: 86  
 Score: 89.09% Conservative: 12  
 Best Local Similarity: 78.18% Mismatches: 12  
 Query Match: 73.72% Indels: 0  
 DB: 19 Gaps: 0

US-09-856-221-4 (1-332) x AAW56572 (1-2516)

QY 3 CAAGCGCAGTATTGCAAAAAAACTATCTGGAACCCCAACAGGCACAACTCAGGCACAG 62  
 DB 2218 GluAlaAlaValLeuGlnIySThrSerLeuIySThrGlnGlnGlnThrGlnSerGln 2237  
 QY 63 CTGGCCTTCCTACAAAGCAAAATTCAGCAATACAGGTTGTATTAAGTGGCTACGTGGCGGA 122  
 DB 2238 LeuAlaPheLeuGlnArgIySPheserAsnGlnAlaLeuIyAsnItrpLeuArgGlyArg 2257  
 QY 123 TTGGCGGCTATTTATATCAGTTTATGACTTGGCTGTTCCCTGTGTTTGTATGGCTGAA 182  
 DB 2258 LeuAlaAlaIleIyPheGlnPheIyAspLeuAlaValAlaArGcysLeuMetAlaGlu 2277  
 QY 183 CAACCTTACCAGTATGAATTTGAACGATAAAGCTGACGCTTCATTAGCCCGGCTCGG 242  
 DB 2278 GlnAlaIyArGrpGluLeuAsnAspSerAlaArGpHeIIeIySProGlyAlaItrp 2297  
 QY 243 CATGCCATTATGCTGTTTGTAGCAGGTGAACCTTGTGCTGAATTTGGCAGATG 302  
 DB 2298 GlnGlyThrTyrAlaGlyLeuLeuAlaGlyGluThrLeuMetLeuSerLeuAlaGlnMet 2317  
 QY 303 GAAAAAACTATTTGGAAGAAGATGAACG 332  
 DB 2318 GluAspAlaHisLeuIySArgAspIySArg 2327

RESULT 5  
 AAW72609  
 ID AAW72609 standard; Protein: 2516 AA.  
 XX  
 AC AAW72609;  
 XX  
 DT 04-MAY-2001 (first entry)  
 DE Photorhabdus tcdA toxin.  
 XX  
 KW TcdA; TcdA; insect toxin; plant; insect resistance.  
 XX  
 OS Photorhabdus sp.  
 XX  
 PN W0200111029-A1.  
 XX  
 PD 15-FEB-2001.  
 XX  
 PF 11-AUG-2000; 2000WO-US22237.  
 XX  
 PR 11-AUG-1999; 99US-0148356.  
 XX  
 PA (DOWC ) DOW AGROSCIENCES LLC.  
 XX  
 PI Petell JK, Merlo DJ, Herman RA, Roberts JL, Guo L, Schafer BW;  
 PI Sukhapinda K, Merlo AO;  
 XX  
 WPI: 2001-191536/19.  
 DR N-PSDB; AAF58778.  
 XX  
 PT Novel polynucleotide sequence encoding insect toxins, useful for  
 PT producing transgenic plants having resistance to insects, especially  
 PT corn rootworm  
 XX  
 PS Disclosure; Page 51-61; 106pp; English.

*Not from arg*

XX The present invention provides the protein and coding sequences of  
 CC modified versions of the Photobabidus TcdA and TcdB toxins. These are  
 CC suitable for expression in plants. The toxins are effective against  
 CC insects upon ingestion, and the sequences provided can be used to produce  
 CC transgenic plants with insect resistance. The present sequence is the  
 CC Photobabidus TcdA protein.  
 XX SQ Sequence 2516 AA;

Alignment Scores: 2.75e-50 Length: 2516  
 Pred. No.: 446.00 Matches: 86  
 Score: 89.09% Conservative: 12  
 Best Local Similarity: 78.18% Mismatches: 12  
 Query Match: 73.72% Indels: 0  
 DB: 22 Gaps: 0

US-09-856-221-4 (1-332) x AAB72609 (1-2516)

QY 3 GAAGCGGAGTATTGCAAAAACCTATCTGGAAACCCACAGGACAACTCAGGCACAG 62  
 Db 2218 GluAlaLaValLeuGlnLysThrSerLeuLysThrGlnGlnGlnThrGlnSerGln 2237  
 QY 63 CTGGCTTCCTACAAAGCAAAATTCAGCAATACAGCGTTGTATTAACCTGCTGGCGGA 122  
 Db 2238 LeuAlaPheLeuGlnArgLysPheSerAsnGlnAlaLeuTyraSnrTrpLeuArgGlyArg 2257  
 QY 123 TTGGGGCTATTATTTATCAGTTTATGACCTTGGCTGTTCCCTGCTGTTGATGGCTGAA 182  
 Db 2258 LeuAlaLaIleTyrPheGlnPheTyraSnrTrpLeuAlaValAlaArgCysLeuMetAlaGlu 2277  
 QY 183 CAACTTACCAAGTATGAATTAAGCAATTAAGCTGTACGCTTCATTAAAGCCGCTGCTGG 242  
 Db 2278 GlnAlaTyraArgTrpGluLeuAsnAspSerAlaArgPheLleLysProGlyAlaTrp 2297  
 QY 243 CATGGCATTATGCTGGTTGTAGCAGTGAACCTTGATGCTGAATTTGGCAGAGT 302  
 Db 2298 GlnGlyThrTyraGlyLeuAlaGlyGluThrLeuMetLeuSerLeuAlaGlnMet 2317  
 QY 303 GAAAAAACTATTGGAAAAAGATGAACGG 332  
 Db 2318 GluAspAlaHisLeuLysArgAspLysArg 2327

RESULT 6  
 AAB72611 standard; Protein; 2517 AA.

XX AAB72611;  
 XX AAB72611;  
 XX 04-MAY-2001 (first entry)  
 XX Modified Photobabidus TcdA toxin.  
 XX TcdA; TcdB; insect toxin; plant; insect resistance.  
 XX Photobabidus sp.  
 OS Synthetic.  
 PN WO200111029-A1.  
 XX 15-FEB-2001.  
 XX 11-AUG-2000; 2000WO-US22237.  
 XX 11-AUG-1999; 99US-0148356.  
 XX (DOW) DOW AGROSCIENCES LLC.  
 XX Petell JK, Merlo DJ, Herman RA, Roberts JL, Guo L, Schafer BW;  
 PI Sukhapinda K, Merlo AO;  
 XX WPI; 2001-191536/19.

DR N-PSDB; AAF58780.  
 XX Novel polynucleotide sequence encoding insect toxins, useful for  
 PT producing transgenic plants having resistance to insects, especially  
 PT corn rootworm.  
 XX Claim 1; Page 72-83; 106pp; English.

XX The present invention provides the protein and coding sequences of  
 CC modified versions of the Photobabidus TcdA and TcdB toxins. These are  
 CC suitable for expression in plants. The toxins are effective against  
 CC insects upon ingestion, and the sequences provided can be used to produce  
 CC transgenic plants with insect resistance. The present sequence is the  
 CC modified Photobabidus TcdA protein.  
 XX SQ Sequence 2517 AA;

Alignment Scores: 2.75e-50 Length: 2517  
 Pred. No.: 446.00 Matches: 86  
 Score: 89.09% Conservative: 12  
 Best Local Similarity: 78.18% Mismatches: 12  
 Query Match: 73.72% Indels: 0  
 DB: 22 Gaps: 0

US-09-856-221-4 (1-332) x AAB72611 (1-2517)

QY 3 GAAGCGGAGTATTGCAAAAACCTATCTGGAAACCCACAGGACAACTCAGGCACAG 62  
 Db 2219 GluAlaLaValLeuGlnLysThrSerLeuLysThrGlnGlnGlnThrGlnSerGln 2238  
 QY 63 CTGGCTTCCTACAAAGCAAAATTCAGCAATACAGCGTTGTATTAACCTGCTGGCGGA 122  
 Db 2239 LeuAlaPheLeuGlnArgLysPheSerAsnGlnAlaLeuTyraSnrTrpLeuArgGlyArg 2258  
 QY 123 TTGGGGCTATTATTTATCAGTTTATGACCTTGGCTGTTCCCTGCTGTTGATGGCTGAA 182  
 Db 2259 LeuAlaLaIleTyrPheGlnPheTyraSnrTrpLeuAlaValAlaArgCysLeuMetAlaGlu 2278  
 QY 183 CAACTTACCAAGTATGAATTAAGCAATTAAGCTGTACGCTTCATTAAAGCCGCTGCTGG 242  
 Db 2279 GlnAlaTyraArgTrpGluLeuAsnAspSerAlaArgPheLleLysProGlyAlaTrp 2298  
 QY 243 CATGGCATTATGCTGGTTGTAGCAGTGAACCTTGATGCTGAATTTGGCAGAGT 302  
 Db 2299 GlnGlyThrTyraGlyLeuAlaGlyGluThrLeuMetLeuSerLeuAlaGlnMet 2318  
 QY 303 GAAAAAACTATTGGAAAAAGATGAACGG 332  
 Db 2319 GluAspAlaHisLeuLysArgAspLysArg 2328

RESULT 7  
 AAB72614

ID AAB72614 standard; Protein; 2537 AA.  
 XX AAB72614;  
 XX AAB72614;  
 XX 04-MAY-2001 (first entry)  
 XX TcdA toxin-zein ER signal peptide fusion protein.  
 DE TcdA; TcdB; insect toxin; plant; insect resistance.  
 XX Chimeric - Photobabidus sp.  
 OS Chimeric - Zea mays.  
 XX WO200111029-A1.  
 XX 15-FEB-2001.  
 XX 11-AUG-2000; 2000WO-US22237.  
 XX 11-AUG-1999; 99US-0148356.

XX PA (DOWC ) DOW AGROSCIENCES LLC.  
 XX PI Petell JK, Merlo DJ, Herman RA, Roberts JL, Guo L, Schafer BW;  
 XX PI Sukhapinda K, Merlo AO;  
 XX DR WPI; 2001-191536/19.  
 XX DR N-PSDB; AAF58783.  
 XX PT Novel polynucleotide sequence encoding insect toxins, useful for  
 PT producing transgenic plants having resistance to insects, especially  
 PT corn rootworm  
 XX PS Example 1: Page 94-104; 106pp; English.  
 XX CC The present invention provides the protein and coding sequences of  
 CC modified versions of the Photobabidus tcdA and tcdB toxins. These are  
 CC suitable for expression in plants. The toxins are effective against  
 CC insects upon ingestion, and the sequences provided can be used to produce  
 CC transgenic plants with insect resistance. The present sequence is a  
 CC fusion protein comprising the Photobabidus tcdA toxin and the maize zein  
 CC protein ER signal peptide.  
 XX SQ Sequence 2537 AA;

Alignment Scores:  
 Pred. No.: 2,76e-50 Length: 2537  
 Score: 446.00 Matches: 86  
 Percent Similarity: 89.09% Conservative: 12  
 Best Local Similarity: 78.18% Mismatches: 12  
 Query Match: 73.72% Indels: 0  
 DB: 22  
 US-09-856-221-4 (1-332) x AAF72614 (1-2537)  
 QY 3 GAAGGGCAGTATTGCAAAAACACTATCTGGAACCCACAGGACAACTCAGGCACAG 62  
 Db 2239 GluAlaLavalLeuGlnLysThrSerLeuLysThrGlnGlnGlnThrGlnSerGln 2258  
 QY 63 CTGGCTTCCTACAAAGCAATTCAGCAATACAGGCTGTGTATTAACCTGCTACGCTGGCGA 122  
 Db 2259 LeuAlaPheLeuGlnAArgPheSerAsnGlnAlaLeuTyrAsnTrpLeuAArgGlyArg 2278  
 QY 123 TTGGGGCTATTATTATTCAGCTTTTATGACTGTGGCTGTTCCTGCTGTTCATGGCTGAA 182  
 Db 2279 LeuAlaLailLeuTyrPheGlnPheTyrAspLeuAlaLailAArgCysLeuMetAlaGln 2298  
 QY 183 CAACTTACCATGATGAATTTGAACGATGAAGCTGTACGCTTCATTAAGCCCGCTGCTGG 242  
 Db 2299 GlnAlaTyrAArgTrpGluLeuAsnAspSerSerAlaAArgPheIleLysProGlyAlaTrp 2318  
 QY 243 CATGGCACTTATGCTGTGTTAGCAGGTGAACCTGTGCTCAATTTGGCAGCATG 302  
 Db 2319 GlnGlyThrTyrAlaGlyLeuLeuAlaGlyGluThrLeuMetLeuSerLeuAlaGlnMet 2338  
 QY 303 GAAAAAACTATTTGGAAAAGATGAACGG 332  
 Db 2339 GluAspAlaHisLeuLysAArgAspLysArg 2348

RESULT 8  
 AAY33729  
 ID AAY33729 standard; Protein: 2522 AA.  
 XX AC AAY33729;  
 XX DT 09-NOV-1999 (first entry)  
 XX DE Photobabidus luminescens hph2-encoded insecticidal toxin.  
 XX KW Symbiotic bacterium; nematode; insect; larva; toxin; insecticide.  
 XX OS Photobabidus luminescens.  
 XX

PN W09942589-A2.  
 XX 26-AUG-1999.  
 XX PF 18-FEB-1999; 99WO-EP01015.  
 XX PR 20-JAN-1999; 99US-0116439.  
 XX PR 20-FEB-1998; 98US-0027080.  
 XX (NOVS ) NOVARTIS AG.  
 PA (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.  
 PA Anderson AR, Chen JS, Dunn MM, Hart HP, Kramer VC;  
 PI Morgan MK, Warren GW;  
 XX WPI; 1999-527479/44.  
 DR N-PSDB; AAZ06831.  
 XX PT New nucleic acid from Photobabidus luminescens encoding insecticidal  
 PT toxins, used for making resistant transgenic plants  
 XX Claim 26; Page 133-140; 148pp; English.  
 XX CC This sequence represents the hph2 gene encoded insecticidal  
 CC toxin from Photobabidus luminescens. It is one of three insecticidal  
 CC toxins (AAV33728-Y33730) encoded by open reading frames (orfs) in a 38kb  
 CC fragment of P. luminescens DNA (AAZ06831). The hph2 gene was identified  
 CC using probe #2 which was amplified from P. luminescens genomic DNA using  
 CC PCR primers AAZ06829-206830. P. luminescens is a member of the  
 CC Enterobacteriaceae family and is a symbiotic bacterium of  
 CC nematodes of the genus Heterorhabditis. The nematodes colonise  
 CC insect larvae, kill them, and their offspring feed on the dead  
 CC larvae. However, the insecticidal agents are produced by P.  
 CC luminescens rather than the nematodes. The toxins have activity against  
 CC Lepidopteran insects such as Cabbage Looper (Trichoplusia ni),  
 CC European Corn Borer (Ostrinia nubilalis) and Fall Armyworm  
 CC (Spodoptera frugiperda) and also against Coleopteran insects  
 CC (e.g., Colorado potato Beetle, Leptinotarsa decimlineata). In  
 CC addition the toxins are active against strains resistant to known  
 CC insecticides. The DNA sequence can be used to generate transgenic plants  
 CC of various species that are resistant to economically important insect  
 CC pests and also for recombinant production of the toxins for use as  
 CC insecticides.  
 XX SQ Sequence 2522 AA;

Alignment Scores:  
 Pred. No.: 1.75e-49 Length: 2522  
 Score: 440.00 Matches: 84  
 Percent Similarity: 88.18% Conservative: 13  
 Best Local Similarity: 76.36% Mismatches: 13  
 Query Match: 72.73% Indels: 0  
 DB: 20  
 US-09-856-221-4 (1-332) x AAY33729 (1-2522)  
 QY 3 GAAGGGCAGTATTGCAAAAACACTATCTGGAACCCACAGGACAACTCAGGCACAG 62  
 Db 2224 GluAlaLavalLeuGlnLysThrSerLeuLysThrGlnGlnGlnThrGlnAlaGln 2243  
 QY 63 CTGGCTTCCTACAAAGCAATTCAGCAATACAGGCTGTGTATTAACCTGCTACGCTGGCGA 122  
 Db 2244 LeuAlaPheLeuGlnAArgLysPheSerAsnGlnAlaLeuTyrAsnTrpLeuAArgGlyArg 2263  
 QY 123 TTGGGGCTATTATTATTCAGTTTTATGACTGTGGCTGTTCCTGCTGTTCATGGCTGAA 182  
 Db 2264 LeuAlaLailLeuTyrPheGlnPheTyrAspLeuAlaLailAArgCysLeuMetAlaGln 2283  
 QY 183 CAACTTACCATGATGAATTTGAACGATGAAGCTGTACGCTTCATTAAGCCCGCTGCTGG 242  
 Db 2284 GlnAlaTyrAArgTrpGluIleSerAspAspSerAlaAArgPheIleLysProGlyAlaTrp 2303  
 QY 243 CATGGCACTTATGCTGTGTTTGTAGCAGGTGAACCTTGTGCTGAATTTGGCAGCATG 302

*Net J Biol Sci*

|||||  
 Db 2304 GlnGlyThrTyrAlaGlyLeuLeuAlaGlyGlnThrLeuMetLeuLeuAlaGlnMet 2323  
 QY 303 GAAMAACCTATTGGGAAAGATGACGG 332  
 Db 2324 GluAspAlaHisLeuArgArgAspLysArg 2333  
 RESULT 9  
 AAW18303  
 ID AAW18303 standard; Protein; 573 AA.  
 XX  
 AC AAW18303;  
 DT 30-JAN-1998 (first entry)  
 XX  
 DE Photorhabdus luminescens insect toxin TcbAIIi.  
 KW Insecticide; insect; toxin; pest control; biological control;  
 KW Photorhabdus luminescens; TcbA; Southern corn rootworm;  
 KW Colorado potato beetle; Western corn rootworm; meal worm;  
 KW boll weevil; turf grub; Coleoptera; beet armyworm; black cutworm;  
 KW cabbage looper; codling moth; corn earworm; European corn borer;  
 KW tobacco hornworm; tobacco budworm; Lepidoptera; Hymenoptera;  
 KW Diptera, Dictyoptera; Acarina; Homoptera.  
 XX  
 OS Photorhabdus luminescens strain W-14 (ATCC 55397).  
 XX  
 PN WO9717432-A1.  
 XX  
 PD 15-MAY-1997.  
 XX  
 PF 06-NOV-1996; 96WO-US18003.  
 XX  
 PR 28-AUG-1996; 96US-0705484.  
 PR 06-NOV-1995; 96US-0007255.  
 PR 28-FEB-1996; 96US-0608423.  
 XX  
 PA (WISC ) WISCONSIN ALUMNI RES FOUND.  
 XX  
 PI Blackburn MB, Bowen DJ, Cliche TA, Ensign JC, Fatig R;  
 PI French-Constant RH, Guo L, Hey TD, Merlo DJ, Orr GL;  
 PI Petell J, Roberts JL, Rocheleau TA, Schoonover S;  
 PI Strickland JA;  
 XX  
 DR WPI; 1997-281022/25.  
 DR N-PSDB; AAT68847.  
 XX  
 PT Photorhabdus sp. insecticidal protein toxins and DNA encoding them -  
 PT can be genetically engineered into insect larvae food and plants for  
 PT insect control  
 XX  
 PS Claim 34; Page 229-231; 276pp; English.  
 XX  
 CC This polypeptide comprises the 62.9 kDa TcbAIIi insecticidal toxin  
 CC protein of Photorhabdus luminescens W-14. Its sequence was deduced  
 CC from a genomic DNA clone (AA17871). TcbAIIi is a proteolytic  
 CC cleavage product of TcbA (see AAW17871). Claimed toxins of P.  
 CC luminescens (see AAW17871, AAW17884-89, AAW17899-900, AAW18301-06) can  
 CC be produced by recombinant DNA methods and applied to, or genetically  
 CC engineered into, insect larvae food and plants for insect control.  
 CC The toxins are particularly effective against Southern corn rootworm,  
 CC Colorado potato beetle, Western corn rootworm, meal worm, boll  
 CC weevil and turf grub (Coleoptera), beet armyworm, black cutworm,  
 CC cabbage looper, codling moth, corn earworm, European corn borer,  
 CC tobacco hornworm and tobacco budworm (Lepidoptera), and are also  
 CC active against insects of the orders Hymenoptera, Diptera,  
 CC Dictyoptera, Acarina and Homoptera. (All claimed).  
 XX  
 SQ Sequence 573 AA;

Alignment Scores: 9.83e-47 Length: 573  
 Pred. No.: 418.00 Matches: 78  
 Score:

Percent Similarity: 84.55% Conservative: 15  
 Best Local Similarity: 70.91% Mismatches: 17  
 Query Match: 69.09% Indels: 0  
 DB: 18 Gaps: 0  
 US-09-856-221-4 (1-332) x AAW18303 (1-573)  
 QY 3 GAAGCGCAGTATTGCAAAAATACTCTGGAACCAACAGGACAACTCAGGACAG 62  
 Db 281 GluAlaAlaGlnMetGlnLysGluTyrLeuLysThrGlnGlnAlaGlnAlaGln 300  
 QY 63 CTGCGCTTCTACAAAGCAATTCAGCAATACAGCGTTGTATACCTGCTACGTGGCGGA 122  
 Db 301 LeuThrPheLeuArgSerLysPheSerAsnGlnAlaLeuTyrSerTrpLeuArgGlyArg 320  
 QY 123 TTGCGCGCTATTATTATATCAGTTTATGACTTGGCTGCTTCCCTGTTTGTATGCTGAA 182  
 Db 321 LeuSerGlyIleTyrPheGlnPheTyrAspLeuAlaValSerArgCysLeuMetAlaGlu 340  
 QY 183 CAACACTTACAGTATGAATGAACGATTAAGCTGTACGCTTCAATTAAGCCGCGCTCGG 242  
 Db 341 GlnSerTyrGlnTrpGluAlaAsnAspAsnSerIleSerPheValLysProGlyAlaTrp 360  
 QY 243 CATGGCACTTATGCTGTTTGTAGCAGGTGAAACCTTGTGCTGAATTTGGCAGATG 302  
 Db 361 GlnGlyThrTyrAlaGlyLeuLeuCysGlyGluAlaLeuIleGlnAsnLeuAlaGlnMet 380  
 QY 303 GAAAAAACTATTGGAAGAAAGATGAACGG 332  
 Db 381 GluGluAlaTyrLeuLysTrpGluSerArg 390  
 RESULT 10  
 AAW56559  
 ID AAW56559 standard; Protein; 573 AA.  
 XX  
 AC AAW56559;  
 XX  
 DT 07-AUG-1998 (first entry)  
 XX  
 DE Toxin TcbAIIi, encoded by the tcbA gene from genomic region tcb.  
 KW Photorhabdus luminescens W-14; nematode; symbiotic;  
 KW Heterorhabditis; tca; tcb; tcc; tcd; insecticidal activity; toxin;  
 KW Lepidoptera; Coleoptera; Hymenoptera; Diptera; Dictyoptera; Acarina;  
 KW mealworm; Southern; Western corn rootworm; Colorado potato beetle;  
 KW mealworm; boll weevil; turf grub; beetle armyworm; black cutworm;  
 KW cabbage looper; codling moth; corn earworm; European corn borer;  
 KW Tobacco hornworm; budworm.  
 XX  
 OS Photorhabdus luminescens.  
 XX  
 PN WO9808932-A1.  
 XX  
 PD 05-MAR-1998.  
 XX  
 PF 05-MAY-1997; 97WO-US07657.  
 XX  
 PR 06-NOV-1996; 96WO-US18003.  
 PR 28-AUG-1996; 96US-0705484.  
 PR 06-NOV-1996; 96US-0743699.  
 XX  
 PA (DOWC ) DOWELANCO.  
 PA (WISC ) WISCONSIN ALUMNI RES FOUND.  
 XX  
 PI Blackburn MB, Bowen DJ, Cliche TA, Ensign JC, Fatig R;  
 PI French-Constant RH, Guo L, Hey TD, Merlo DJ, Orr GL;  
 PI Petell J, Roberts JL, Rocheleau TA, Schoonover S;  
 PI Strickland JA, Sukhapinda K;  
 XX  
 DR WPI; 1998-179427/16.  
 DR N-PSDB; AAV29987.  
 XX  
 PT Isolated toxins from Photorhabdus luminescens strains - useful for

PT control of insect pests

PS Claim 34; Pages 261-263; 32lpp; English.

XX The present sequence represents a protein named TcbA111 of the bacterium  
XX Photobacterium luminescens (W-14). This is a symbiotic bacterium of the  
CC nematodes of the Heterorhabditis genus. The bacterium has at least 4  
CC distinct genomic regions, tca, tcb, tcc, and tcd. Peptide products are  
CC produced from these regions that are associated with insecticidal  
CC activity. The native toxins are secreted proteins. The proteins are  
CC toxic to insects upon exposure and especially when ingested. The  
CC nucleic acid sequence can be used to produce transgenic plants,  
CC baculoviruses or microbial hosts for toxin production. They can be used  
CC to control insects pests from the Lepidoptera, Coleoptera, Hymenoptera,  
CC Diptera, Dictyoptera, Acarina or Homoptera orders, especially the  
CC Southern or Western corn rootworm, Colorado potato beetle, mealworm,  
CC boll weevil, turf grub, beetle armyworm, black cutworm, cabbage looper,  
CC codling moth, corn earworm, European corn borer or tobacco hornworm  
CC or budworm.

XX SQ Sequence 573 AA;

Alignment Scores:

Pred. No.: 9,83e-47 Length: 573  
Score: 418.00 Matches: 78  
Percent Similarity: 84.55% Conservative: 15  
Best Local Similarity: 70.91% Mismatches: 17  
Query Match: 69.09% Indels: 0  
DB: 19 Gaps: 0

US-09-856-221-4 (1-332) x AAW56557 (1-573)

QY 3 GAACGGCAGTATTGCAAAACAACTATCTGGAACCCAGCCAACTCAGGCACAG 62  
Db 281 GluAlaAlaGluMetGlnLysGluTyrLeuTyrThrGlnGlnAlaGlnAlaGln 300  
QY 63 CTGGCCTTCCTCAACAGCAATTCAGCAATACAGCGTTGTATTAACCTGCTACGCGCGA 122  
Db 301 LeuThrPheLeuArgSerLysPheSerAsnGlnAlaLeuTyrSerTrpLeuArgGlyArg 320  
QY 123 TTGCGCGCTATTATATATACATGTTTATGACTGGCTGTTCCCTGCTGTTGCTGCTGAA 182  
Db 321 LeuSerGlyLeuTyrPheGlnPheTyrAspLeuAlaValSerArgCysLeuMetAlaGlu 340  
QY 183 CAACCTACAGTATGAATTGAACGATAAAGCTGTACGCTTCATTAAAGCCGCTGCTGG 242  
Db 341 GlnSerTyrGlnTrpGluAlaAsnAspAsnSerLysPheValLysProGlyAlaTrp 360  
QY 243 CATGGCCTTATGCTGTTTGTAGCAGTGAACTGCTGCTGCTGCTGCTGCTGCTGCTG 302  
Db 361 GlnGlyThrTyrAlaGlyLeuLeuCysGlyGluAlaLeuLeuGlnAsnLeuAlaGlnMet 380  
QY 303 GAAAAAACTATTGGAAGAAAGATGAACG 332  
Db 381 GluGluAlaTyrLeuTyrTrpGluSerArg 390

RESULT 11

AAW56557

ID AAW56557 standard; Protein; 2504 AA.

XX AAW56557;

XX 07-AUG-1998 (first entry)

XX Toxin TcbA, encoded by the tcbA gene from genomic region tcb.

DE Photobacterium luminescens W-14; nematode; symbiotic;  
XX Heterorhabditis; tca; tcb; tcc; tcd; insecticidal activity; toxin;  
KW Lepidoptera; Coleoptera; Hymenoptera; Diptera; Dictyoptera; Acarina;  
KW Homoptera; Southern; Western corn rootworm; Colorado potato beetle;  
KW mealworm; boll weevil; turf grub; beetle armyworm; black cutworm;  
KW cabbage looper; codling moth; corn earworm; European corn borer;  
KW Tobacco hornworm; budworm.

XX Photobacterium luminescens.

XX WO9808932-A1.

XX 05-MAR-1998.

XX 05-MAY-1997; 97WO-US07657.

XX 06-NOV-1996; 96WO-US18003.

XX 28-AUG-1996; 96US-0705484.

XX 06-NOV-1996; 96US-0743699.

XX (DOWC) DOWELANCO.

PA (WISC) WISCONSIN ALUMNI RES FOUND.

XX Blackburn MB, Bowen DJ, Ciche TA, Ensign JC, Fatig R;

PI Ffrench-Constant RH, Guo L, Hey TD, Merlo DJ, Orr GL;

PI Petell J, Roberts JL, Rocheleau TA, Schoonover S;

PI Strickland JA, Sukhapinda K;

XX WPI; 1998-179427/16.

DR N-PSDB; AAV29985.

XX Isolated toxins from Photobacterium luminescens strains - useful for control of insect pests

XX Claim 34; Pages 163-169; 32lpp; English.

XX The present sequence represents a protein named TcbA of the bacterium  
CC Photobacterium luminescens (W-14). This is a symbiotic bacterium of the  
CC nematodes of the Heterorhabditis genus. The bacterium has at least 4  
CC distinct genomic regions, tca, tcb, tcc, and tcd. Peptide products are  
CC produced from these regions that are associated with insecticidal  
CC activity. The native toxins are secreted proteins. The proteins are  
CC toxic to insects upon exposure and especially when ingested. The  
CC nucleic acid sequence can be used to produce transgenic plants,  
CC baculoviruses or microbial hosts for toxin production. They can be used  
CC to control insects pests from the Lepidoptera, Coleoptera, Hymenoptera,  
CC Diptera, Dictyoptera, Acarina or Homoptera orders, especially the  
CC Southern or Western corn rootworm, Colorado potato beetle, mealworm,  
CC boll weevil, turf grub, beetle armyworm, black cutworm, cabbage looper,  
CC codling moth, corn earworm, European corn borer or tobacco hornworm  
CC or budworm.

XX SQ Sequence 2504 AA;

Alignment Scores:

Pred. No.: 1.53e-46 Length: 2504  
Score: 418.00 Matches: 78  
Percent Similarity: 84.55% Conservative: 15  
Best Local Similarity: 70.91% Mismatches: 17  
Query Match: 69.09% Indels: 0  
DB: 19 Gaps: 0

US-09-856-221-4 (1-332) x AAW56557 (1-2504)

QY 3 GAACGGCAGTATTGCAAAACAACTATCTGGAACCCAGCCAACTCAGGCACAG 62

Db 2212 GluAlaAlaGluMetGlnLysGluTyrLeuTyrThrGlnGlnAlaGlnAlaGln 2231

QY 63 CTGGCCTTCCTCAACAGCAATTCAGCAATACAGCGTTGTATTAACCTGCTACGCGCGA 122

Db 2232 LeuThrPheLeuArgSerLysPheSerAsnGlnAlaLeuTyrSerTrpLeuArgGlyArg 2251

QY 123 TTGCGCGCTATTATATATACATGTTTATGACTGGCTGTTCCCTGCTGTTGATGGCTGAA 182

Db 2252 LeuSerGlyLeuTyrPheGlnPheTyrAspLeuAlaValSerArgCysLeuMetAlaGlu 2271

QY 183 CAACCTTACAGTATGAATTGAACGATAAAGCTGTACGCTTCATTAAAGCCGCTGCTGG 242

Db 2272 GlnSerTyrGlnTrpGluAlaAsnAspAsnSerLysPheValLysProGlyAlaTrp 2291

QY 243 CATGCCACTTATGCTGTTTGTAGCAGTGAAACCTTGATGCTGAATTTGGCAGCATG 302  
 Db 2292 GlnGlyThrYrAlaGlyLeuLeuCysGlyGluAlaLeuLeuGlnAsnLeuAlaGlnMet 2311  
 QY 303 GAAAAAATTTTGGAAAAAGATGAACGG 332  
 Db 2312 GluGluAlaTyrLeuLysTrpGluSerArg 2321  
 RESULT 12  
 AAB72610  
 ID AAB72610 standard; Protein; 2504 AA.  
 XX AAB72610;  
 XX  
 DT 04-MAY-2001 (first entry)  
 XX  
 DE Photorhabdus tcbA toxin.  
 XX  
 KW TcdA; TcBA; insect toxin; plant; insect resistance.  
 XX  
 OS Photorhabdus sp.  
 XX  
 PN WO200111029-A1.  
 XX  
 PD 15-FEB-2001.  
 XX  
 PF 11-AUG-2000; 2000WO-US22237.  
 XX  
 PR 11-AUG-1999; 99US-0148356.  
 XX  
 PA (DOWC ) DOW AGROSCIENCES LLC.  
 XX  
 PI Petell JK, Merlo DJ, Herman RA, Roberts JL, Guo L, Schafer BW;  
 PI Sukhapinda K, Merlo AO;  
 XX  
 DR WPI; 2001-191536/19.  
 DR N-PSDB; AAF58779.  
 XX  
 PT Novel polynucleotide sequence encoding insect toxins, useful for  
 PT producing transgenic plants having resistance to insects, especially  
 PT corn rootworm -  
 XX  
 PS Disclosure; Page 62-72; 106pp; English.  
 XX  
 CC The present invention provides the protein and coding sequences of  
 CC modified versions of the Photorhabdus tcbA and tcbA toxins. These are  
 CC suitable for expression in plants. The toxins are effective against  
 CC insects upon ingestion, and the sequences provided can be used to produce  
 CC transgenic plants with insect resistance. The present sequence is the  
 CC Photorhabdus TcBA protein.  
 XX  
 SQ Sequence 2504 AA;  
 Alignment Scores:  
 Pred. No.: 1.53e-46 Length: 2504  
 Score: 418.00 Matches: 78  
 Percent similarity: 84.55% Conservative: 15  
 Best Local Similarity: 70.91% Mismatches: 17  
 Query Match: 69.09% Indels: 0  
 DB: 22 Gaps: 0  
 US-09-856-221-4 (1-332) x AAB72610 (1-2504)  
 QY 3 GAAGCGCAGTATTGGCAAAAATACTATCTGGAAACCCACAGGCACAACTCAGGCACAG 62  
 Db 2212 GluAlaLaGluMetGlnLysGluTyrLeuLysThrGlnGlnAlaGlnAlaGln 2231  
 QY 63 CTGGCTTCCTACAAAGCAATTCAGCAATACAGGTTGTATTAACCTGCTACGTGGCGA 122  
 Db 2232 LeuThrPheLeuArgSerLysPheSerAsnGlnAlaLeuTyrSerTrpLeuArgGlyAla 2251  
 QY 123 TTGGCGGCTATTATTATACAGTTTATGACTTGGCTGTTCCTGCTGTGTGATGGCTGAA 182  
 Db 2213 GluAlaLaGluMetGlnLysGluTyrLeuLysThrGlnGlnAlaGlnAlaGln 2232

Db 2252 LeuSerGlyIleTyrPheGlnPheTyrAspLeuAlaValSerArgCysLeuMetAlaGlu 2271  
 QY 193 CAACATTACAGTATGATGAATTAAGCATTAAGCTGCTTCAATTAAGCCGCGTCCCTGG 242  
 Db 2272 GlnSerTyrGlnTrpGluAlaAsnAspSerIleSerPheValLysProGlyAlaTrp 2291  
 QY 243 CATGCCACTTATGCTGTTTGTAGCAGTGAAACCTTGATGCTGAATTTGGCAGCATG 302  
 Db 2292 GlnGlyThrYrAlaGlyLeuLeuCysGlyGluAlaLeuLeuGlnAsnLeuAlaGlnMet 2311  
 QY 303 GAAAAAATTTTGGAAAAAGATGAACGG 332  
 Db 2312 GluGluAlaTyrLeuLysTrpGluSerArg 2321  
 RESULT 13  
 AAB72612  
 ID AAB72612 standard; Protein; 2505 AA.  
 XX AAB72612;  
 XX  
 DT 04-MAY-2001 (first entry)  
 XX  
 DE Modified Photorhabdus tcbA toxin.  
 XX  
 KW TcdA; TcBA; insect toxin; plant; insect resistance.  
 XX  
 OS Photorhabdus sp.  
 XX  
 PN WO200111029-A1.  
 XX  
 PD 15-FEB-2001.  
 XX  
 PF 11-AUG-2000; 2000WO-US22237.  
 XX  
 PR 11-AUG-1999; 99US-0148356.  
 XX  
 PA (DOWC ) DOW AGROSCIENCES LLC.  
 XX  
 PI Petell JK, Merlo DJ, Herman RA, Roberts JL, Guo L, Schafer BW;  
 PI Sukhapinda K, Merlo AO;  
 XX  
 DR WPI; 2001-191536/19.  
 DR N-PSDB; AAF58781.  
 XX  
 PT Novel polynucleotide sequence encoding insect toxins, useful for  
 PT producing transgenic plants having resistance to insects, especially  
 PT corn rootworm -  
 XX  
 PS Claim 1; Page 83-93; 106pp; English.  
 XX  
 CC The present invention provides the protein and coding sequences of  
 CC modified versions of the Photorhabdus tcbA and tcbA toxins. These are  
 CC suitable for expression in plants. The toxins are effective against  
 CC insects upon ingestion, and the sequences provided can be used to produce  
 CC transgenic plants with insect resistance. The present sequence is the  
 CC modified Photorhabdus TcBA protein.  
 XX  
 SQ Sequence 2505 AA;  
 Alignment Scores:  
 Pred. No.: 1.53e-46 Length: 2505  
 Score: 418.00 Matches: 78  
 Percent similarity: 84.55% Conservative: 15  
 Best Local Similarity: 70.91% Mismatches: 17  
 Query Match: 69.09% Indels: 0  
 DB: 22 Gaps: 0  
 US-09-856-221-4 (1-332) x AAB72612 (1-2505)  
 QY 3 GAAGCGCAGTATTGGCAAAAATACTATCTGGAAACCCACAGGCACAACTCAGGCACAG 62  
 Db 2213 GluAlaLaGluMetGlnLysGluTyrLeuLysThrGlnGlnAlaGlnAlaGln 2232

QY 63 CTGGCTTCCTACAAAGCAATTCAGCAATACAGCGTTGTATACCTGGCTACGTGGCGCA 122  
 Db 2233 LeuThrPheLeuArgSerLysPheSerAsnGlnAlaLeuTyrSerTrpLeuArgGlyArg 2252  
 QY 123 TTGGCGGCTATTATATACAGTTTATGACTTGGCTGGTTCCTCGTGTGTTGATGGCTGAA 182  
 Db 2253 LeuSerGlyIleTyrPheGlnPheTyrAspLeuAlaValSerArgCysLeuMetAlaGlu 2272  
 QY 183 CAACTTACCAGTATGATTAACAGGATAAAGCTTACGCTTCATTAAGCCCGCTGG 242  
 Db 2273 GlnSerTyrGlnTrpGluAlaAsnAspAsnSerIleSerPheValLysProGlyAlaIleTrp 2292  
 QY 243 CATGGCACTTATGCTGGTTTGTAGCAGGTGAACCTTCATGCTGCTGAATTTGGCACAGATG 302  
 Db 2293 GlnGlyThrTyrAlaGlyLeuLeuCysGlyGluAlaLeuIleGlnAsnLeuAlaGlnMet 2312  
 QY 303 GAAAAAACTATTGGAAAAAGATGACGG 332  
 Db 2313 GluGluAlaTyrLeuLysTrpGluSerArg 2322  
 RESULT 14  
 AAW17871  
 ID AAW17871 standard; Protein; 2504 AA.  
 XX  
 AC AAW17871;  
 XX  
 DT 29-JAN-1998 (first entry)  
 XX  
 DE Photorhabdus luminescens insect toxin protein TcBa.  
 XX  
 KW Insecticide; insect; toxin; pest control; biological control;  
 KW Photorhabdus luminescens; TcBa; Southern corn rootworm;  
 KW Colorado potato beetle; Western corn rootworm; meal worm;  
 KW boll weevil; turf grub; Coleoptera; beet armyworm; black cutworm;  
 KW cabbage looper; codling moth; corn earworm; European corn borer;  
 KW tobacco hornworm; tobacco budworm; Lepidoptera; Hymenoptera;  
 KW Diptera, Dictyoptera; Acarina; Homoptera.  
 XX  
 OS Photorhabdus luminescens strain W-14 (ATCC 55397).  
 XX  
 PN WO917432-Al.  
 XX  
 PD 15-MAY-1997.  
 XX  
 PF 06-NOV-1996; 96WO-US18003.  
 XX  
 PR 28-AUG-1996; 96US-0705484.  
 PR 06-NOV-1995; 95US-0007255.  
 PR 28-FEB-1996; 96US-0608423.  
 XX  
 PA (WISC ) WISCONSIN ALUMNI RES FOUND.  
 XX  
 PI Blackburn MB, Bowen DJ, Cliche TA, Ensign JC, Fatig R;  
 PI French-Constant RH, Guo L, Hey TD, Merlo DJ, Orr GL;  
 PI Petell J, Roberts JL, Rocheleau TA, Schoonover S;  
 PI Strickland JA;  
 XX  
 DR WPI: 1997-281022/25.  
 DR N-PSDB; AAT68836.  
 XX  
 PT Photorhabdus sp. insecticidal protein toxins and DNA encoding them -  
 PT can be genetically engineered into insect larvae food and plants for  
 PT insect control  
 XX  
 PS Claim 8; Page 119-129; 276pp; English.  
 XX  
 CC This polypeptide comprises a specifically claimed 280.6 kDa  
 CC insecticidal toxin, TcBa, of Photorhabdus luminescens that is  
 CC proteolytically cleaved into 207.6 kDa (see AAW18302) and 62.9 kDa  
 CC polypeptides (see AAW18303). The TcBa polypeptide can be expressed  
 CC in host cells using a gene (see AAT68836) isolated from a genomic  
 CC library. Claimed toxin proteins of P. luminescens (see AAW17871,  
 CC

CC AAW17884-89, AAW17899-900, AAW18301-06) can be applied to, or  
 CC genetically engineered into, insect larvae food and plants for insect  
 CC control. The Photorhabdus toxins are particularly effective against  
 CC Southern corn rootworm, Colorado potato beetle, Western corn rootworm,  
 CC meal worm, boll weevil and turf grub (Coleoptera), beet armyworm, black  
 CC cutworm, cabbage looper, codling moth, corn earworm, European corn  
 CC borer, tobacco hornworm and tobacco budworm (Lepidoptera), and  
 CC are also active against insects of the orders Hymenoptera, Diptera  
 CC Dictyoptera, Acarina and Homoptera. (All claimed).  
 XX  
 SQ Sequence 2504 AA;  
 Alignment Scores:  
 Pred. No.: 1,79e-45 Length: 2504  
 Score: 410.00 Matches: 77  
 Percent Similarity: 83.64% Conservatives: 15  
 Best Local Similarity: 70.00% Mismatches: 18  
 Query Match: 67.77% Indels: 0  
 DB: 18 Gaps: 0  
 US-09-856-221-4 (1-332) x AAW17871 (1-2504)  
 QY 3 GAAAGCGCAGTATTGCAAAAACTATCTGGAAACCCACAGGCACAACTCAGGCACAG 62  
 Db 2212 GluAlaAlaGluMetGlnLysGluTyrLeuLysThrGlnGlnAlaGlnAlaGln 2231  
 QY 63 CTGGCCTTCCTACAAAGCAATTCAGCAATACAGCGTTGTATACCTGGCTGGCGCA 122  
 Db 2232 LeuThrPheLeuArgSerLysPheSerAsnGlnAlaLeuTyrSerTrpLeuArgGlyArg 2251  
 QY 123 TTGGCGGCTATTATATACAGTTTATGACTTGGCTGTTCCTGTGTTTATGATGGCTGAA 182  
 Db 2252 LeuSerGlyIleTyrPheGlnPheTyrAspLeuAlaValSerArgCysLeuMetAlaGlu 2271  
 QY 183 CAACTTACCAGTATGATTAACAGGATAAAGCTTCATGCTGCTGAATTTGGCACAGATG 302  
 Db 2272 GlnSerTyrGlnTrpGluAlaAsnAspAsnSerIleSerPheValLysProGlyAlaIleTrp 2291  
 QY 243 CATGGCACTTATGCTGGTTTGTAGCAGGTGAACCTTCATGCTGCTGAATTTGGCACAGATG 302  
 Db 2292 GlnGlyThrTyrAlaGlyLeuLeuCysGlyGluAlaLeuIleGlnMetLeuAlaGlnMet 2311  
 QY 303 GAAAAAACTATTGGAAAAAGATGACGG 332  
 Db 2312 GluGluAlaTyrLeuLysTrpGluSerArg 2321  
 RESULT 15  
 AAY97694  
 ID AAY97694 standard; Protein; 2376 AA.  
 XX  
 AC AAY97694;  
 XX  
 DT 19-JUN-2001 (first entry)  
 XX  
 DE SepA protein encoded by Serratia insecticidal protein complex gene.  
 XX  
 KW Insecticidal protein complex; amber disease; insect; Coleoptera;  
 KW pesticide; SepA protein.  
 XX  
 OS Serratia sp.  
 XX  
 PN WO200116305-A2.  
 XX  
 PD 08-MAR-2001.  
 XX  
 PF 04-SEP-2000; 2000WO-NZ00174.  
 XX  
 PR 02-SEP-1999; 99NZ-0337610.  
 PA (NZPA-) NEW ZEALAND PASTORAL AGRIC RES INST LTD.  
 XX  
 PI Glare TR, Hurst MRH, Jackson TA;  
 XX







GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model  
Run on: January 15, 2003, 15:39:33 ; Search time 41.5673 Seconds  
(without alignments)  
2878.683 Million cell updates/sec

Title: US-09-856-221-1  
Perfect score: 793  
Sequence: 1 tgtcagcagcactcacttat.....gtgaccacactgatacagtt 449

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Zgapop 6.0, Zgapext 7.0  
Delep 6.0, Delext 7.0

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameter--

Minimum DB seq length: 0  
Maximum DB seq length: 200000

Post-processing: Minimum Match:  
Maximum Match:  
Listing first

Command line parameters:  
-O=/cn2.1/USPTO.spool/US098562  
-DB=A\_Geneseq\_101002-QPNT-fasta  
-LOPEXT=0 -UNITS=bits -THRT=1  
-LIST=45 -DOCALIGN=200 -THRT=SCOR  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext  
-USER=US09856221\_CCGN\_1.1\_60.ern  
-NO\_XLPXY -NO\_MMAP -LARGEQUERY -N  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A\_Geneseq\_101002.\*  
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	434.5	54.8	2516	18	AAW17899
2	434.5	54.8	2516	19	AAW56572
3	434.5	54.8	2516	22	AAW2609
4	434.5	54.8	2517	22	AAW2611
5	434.5	54.8	2537	22	AAW2614
6	428.5	54.0	2522	20	AAW33729
7	416.5	52.5	2376	22	AAW97694
8	415.5	52.3	2504	19	AAW56557
9	415.5	52.3	2504	22	AAW2610
10	412.5	52.0	2505	22	AAW2612
11	410.5	51.7	2504	18	AAW17871
12	354.5	44.6	579	18	AAW18301
13	354.5	44.6	579	19	AAW56574
14	353.5	44.5	573	18	AAW18303
15	353.5	44.5	573	19	AAW56559
16	227.5	28.6	1189	18	AAW17884
17	227.5	28.6	1189	19	AAW56543
18	208.5	26.2	1565	19	AAW56568
19	201.5	25.3	1562	18	AAW17886
20	201.5	25.3	1562	19	AAW56545
21	190.5	24.0	1565	18	AAW18305
22	93.5	11.7	1849	18	AAW17900
23	93.5	11.7	1849	19	AAW56573
24	84.5	10.7	568	22	AAW27972
25	84.5	10.7	568	23	AAW90550
26	81.5	10.2	578	22	AAW71138
27	80.5	10.2	446	22	AAW63108
28	78.5	9.9	662	22	AAW00477
29	77.5	9.7	3201	22	AAW62899
30	76.5	9.6	841	22	ABG19955
31	76.5	9.6	1014	23	AAW96745
32	76.5	9.6	1447	22	AAW86975
33	76.5	9.6	2799	22	AAW81867
34	74.5	9.3	1434	22	AAW83976
35	73.5	9.3	338	23	AAW77921
36	73.5	9.3	371	20	AAW73369
37	73.5	9.3	716	22	AAW28268
38	73.5	9.2	149	22	AAW64213
39	72.5	9.1	805	22	AAW62304
40	72.5	9.1	926	22	AAW62874
41	72.5	9.1	4498	22	AAW58595
42	72.5	9.1	1905	22	AAW59243
43	72.5	9.1	1905	22	AAW38925
44	71.5	9.0	399	21	AAW34742
45	71.5	9.0	694	20	AAW17829

## ALIGNMENTS

RESULT 1  
AAW17899  
ID AAW17899 standard; Protein; 2516 AA.

XX  
AC AAW17899;  
XX  
DF 29-JAN-1998 (first entry)  
XX  
DE Photorhabdus luminescens insect toxin TcdA.  
XX  
DE Insecticide; insect; toxin; pest control; biological control;  
KW Photorhabdus luminescens; TcdA; Southern corn rootworm;  
KW Colorado potato beetle; Western corn rootworm; meal worm;  
KW boll weevil; turf grub; Coleoptera; beet armyworm; black cutworm;  
KW cabbage looper; codling moth; corn earworm; European corn borer;  
KW tobacco hornworm; tobacco budworm; Lepidoptera; Hymenoptera;  
KW Diptera, Dictyoptera; Acarina; Homoptera.  
XX  
OS Photorhabdus luminescens strain W-14 (ATCC 55397).

XX FH Key Location/Qualifiers  
 FT Protein 1..2516  
 FT /label= TcdA  
 FT Protein 89..1937  
 FT /label= TcdAii  
 FT Peptide 89..100  
 FT /note= "S2 N-terminus (Claim 30)"  
 FT Peptide 284..299  
 FT /note= "tryptic peptide (Claim 30)"  
 FT Peptide 554..563  
 FT /note= "tryptic peptide (Claim 30)"  
 FT Peptide 1080..1092  
 FT /note= "Isolated N-terminal peptide (Claim 30)"  
 FT Peptide 1385..1400  
 FT /note= "tryptic peptide (Claim 30)"  
 FT Peptide 1478..1497  
 FT /note= "tryptic peptide (Claim 30)"  
 FT Peptide 1620..1642  
 FT /note= "claimed peptide (Claim 30)"  
 FT Peptide 1938..1948  
 FT /note= "N-terminal peptide (Claim 30)"  
 FT Protein 1938..2516  
 FT /label= TcdAii  
 FT Peptide 2327..2345  
 FT /note= "tryptic peptide (Claim 30)"  
 FT Peptide 2398..2408  
 FT /note= "tryptic peptide (Claim 30)"  
 XX WO9717432-A1.  
 XX  
 XX 15-MAY-1997.  
 XX 06-NOV-1996; 96WO-US18003.  
 XX 28-AUG-1996; 96US-0705484.  
 XX 06-NOV-1995; 95US-0007255.  
 XX 28-FEB-1996; 96US-0608423.  
 XX  
 XX (WISC ) WISCONSIN ALUMNI RES FOUND.  
 XX  
 XX Blackburn MB, Bowen DJ, Cliche TA, Ensign JC, Fatig R;  
 PI French-Constant RH, Guo L, Hey TD, Merlo DJ, Orr GL;  
 PI Petell J, Roberts JL, Rocheleau TA, Schoonover S;  
 PI Strickland JA;  
 XX  
 XX WPI; 1997-281022/25.  
 XX N-PSDB; AAT68843.  
 XX  
 XX Photorhabdus sp. insecticidal protein toxins and DNA encoding them -  
 PT can be genetically engineered into insect larvae food and plants for  
 PT insect control  
 XX  
 XX Claim 34; Page 185-194; 276pp; English.  
 XX  
 XX This polypeptide comprises the 282.9 kDa TcdA insecticidal toxin  
 CC protein of Photorhabdus luminescens W-14. Its sequence was deduced  
 CC from a genomic DNA clone (AAT68843) and includes N-terminal and  
 CC tryptic peptide sequences obtained from the isolated protein. The  
 CC 282.9 kDa insect toxin is proteolytically processed into TcdAii and  
 CC TcdAiii components (see AAW17900 and AAW18301). Claimed toxins of P.  
 CC luminescens (see AAW17871, AAW17884-89, AAW17899-900, AAW18301-06) can  
 CC be produced by recombinant DNA methods and applied to, or genetically  
 CC engineered into, insect larvae food and plants for insect control.  
 CC The toxins are particularly effective against Southern corn rootworm,  
 CC Colorado potato beetle, Western corn rootworm, meal worm, boll weevil  
 CC and turf grub (Coleoptera), beet armyworm, black cutworm, cabbage  
 CC looper, codling moth, corn earworm, European corn borer, tobacco  
 CC hornworm and tobacco budworm (Lepidoptera), and are also active  
 CC against insects of the orders Hymenoptera, Diptera, Dictyoptera,  
 CC Acarina and Homoptera. (All claimed).  
 XX  
 XX Sequence 2516 AA;

Alignment Scores:  
 Pred. NO.: 9,52e-46 Length: 2516  
 Score: 434.50 Matches: 87  
 Percent Similarity: 72.30% Conservative: 20  
 Best Local Similarity: 58.78% Mismatches: 38  
 Query Match: 54.79% Indels: 3  
 DB: 18 Gaps: 1

US-09-856-221-1 (1-449) x AAW17899 (1-2516)

QY 13 CACCTTATGGCGATAACCTTATTTTCATTTGGATTAACGATTGGTCAGAACCCGGTTTA 72  
 |||||:||||| ||||| |||||:||||| |||||  
 Db 1888 HisLeuLeuGlyAspLysProTyrLeuProLeuSerThrThrTrpSerAspProArgLeu 1907  
 QY 73 GAAGAAGCCGCCAGTCAAAACCATTCGTGATCATTAACGATTAATAATGCGCAACTGCGT 132  
 :||| ||||| |||||: |||||: |||||  
 Db 1908 AspArgAlaAlaAspIleThrThrGlnAsnAlaHisAspSerAlaIleValAlaLeuArg 1927  
 :||| ||||| |||||: |||||: |||||  
 QY 133 CAACGCGCGCCTTGGCGCGGAAA-----CGTACTGCAAAATTCGTTAACCGCTTTG 183  
 ||||| ||||| |||||: |||||: |||||  
 Db 1928 GlnAsnIleProThrProAlaProLeuSerLeuArgSerAlaAsnThrLeuThrAspLeu 1947  
 QY 184 TTCCTTCCTCAGATAAACAATAAACTGCAAAAGTTACTGGCAGACGTTAGCACACGCGCTA 243  
 ||||| |||||: |||||: |||||: |||||: |||||  
 Db 1948 PheLeuProGlnIleAsnGluValMetMetAsnTyrTrpGlnThrLeuAlaGlnArgVal 1967  
 QY 244 TATAACTTACGTCAATCTGCAATTTGATGTCAGCGGTTGTCATACCATCTATCGG 303  
 ||||| |||||: |||||: |||||: |||||: |||||  
 Db 1968 TyrAsnLeuArgHisAsnLeuSerIleAspGlyGlnProLeuTyrLeuProIleTyrAla 1987  
 QY 304 ACACGACGAGATCCGTCCTACTGCTTAGTGTCTGCTCCGCTCACCGCTCACAAAGCGGAGG 363  
 ||||| ||||| |||||: |||||: |||||  
 Db 1988 ThrProAlaAspProLysAlaLeuLeuSerAlaAlaValAlaThrSerGlnGlyGly 2007  
 QY 364 GATTTCCTCGGACAGTAATGCGGATGACGTTTTCGATTTTCCGATTATTCGAAAATGCCAAG 423  
 ||||| |||||: |||||: |||||: |||||: |||||  
 Db 2008 LysLeuProGluSerPheMetSerLeuIrrpArgPheProHisMetLeuGluAsnAlaArg 2027

QY 424 TGGGAGTGACCACTGATACAG 447

Db 2028 GlyMetValSerGlnLeuThrGln 2035  
 |||||:||||| |||||

RESULT 2  
 AAW56572  
 ID AAW56572 standard; Protein; 2516 AA.  
 XX  
 AC AAW56572;  
 XX  
 DT 07-AUG-1998 (first entry)  
 XX  
 DE Toxin TcdA, encoded by the tcdA gene from genomic region tcd.  
 XX  
 KW Photorhabdus luminescens W-14; nematode; symbiotic;  
 KW Heterorhabditis; tca; tcb; tcc; tcd; insecticidal activity; toxin;  
 KW Lepidoptera; Coleoptera; Hymenoptera; Diptera; Dictyoptera; Acarina;  
 KW Homoptera; Southern; Western corn rootworm; Colorado potato beetle;  
 KW mealworm; boll weevil; turf grub; beetle armyworm; black cutworm;  
 KW cabbage looper; codling moth; corn earworm; European corn borer;  
 KW Tobacco hornworm; budworm.  
 XX  
 OS Photorhabdus luminescens.  
 XX  
 PN WO9808932-A1.  
 XX  
 XX 05-MAR-1998.  
 PD  
 XX 05-MAY-1997; 97WO-US07657.  
 PF  
 XX 06-NOV-1996; 96WO-US18003.  
 PR  
 XX 28-AUG-1996; 96US-0705484.  
 PR  
 XX 06-NOV-1996; 96US-0743699.  
 XX

PA (DOWC ) DOWELANCO.  
 PA (WISC ) WISCONSIN ALUMNI RES FOUND.  
 XX Blackburn MB, Bowen DJ, Cliche TA, Ensign JC, Fatig R;  
 PI French-Constant RH, Guo L, Hey TD, Merlo DJ, Orr GL;  
 PI Petell J, Roberts JL, Rocheleau TA, Schoonover S;  
 PI Strickland JA, Sukhapinda K;  
 XX WPI: 1998-179427/16.  
 DR N-PSDB; AAV2992B.  
 XX Isolated toxins from Photorhabdus luminescens strains - useful for  
 PT control of insect pests  
 XX Claim 34; Pages 224-231; 321pp; English.  
 XX The present sequence represents a protein named TcdA of the bacterium  
 CC Photorhabdus luminescens (W-14). This is a symbiotic bacterium of the  
 CC nematodes of the Heterorhabditis genus. The bacterium has at least 4  
 CC distinct genomic regions, tca, tcb, tcc, and tcd. Peptide products are  
 CC produced from these regions that are associated with insecticidal  
 CC activity. The native toxins are secreted proteins. The proteins are  
 CC toxic to insects upon exposure and especially when ingested. The  
 CC nucleic acid sequence can be used to produce transgenic plants,  
 CC baculoviruses or microbial hosts for toxin production. They can be used  
 CC to control insects pests from the Lepidoptera, Coleoptera, Hymenoptera,  
 CC Diptera, Dictyoptera, Acarina or Homoptera orders, especially the  
 CC Southern or Western corn rootworm, Colorado potato beetle, mealworm,  
 CC boll weevil, turf grub, beetle armyworm, black cutworm, cabbage looper,  
 CC codling moth, corn earworm, European corn borer or tobacco hornworm  
 CC or budworm.  
 XX Sequence 2516 AA:

Alignment Scores:  
 Pred. No.: 9,52e-46 Length: 2516  
 Score: 434.50 Matches: 87  
 Percent Similarity: 72.30% Conservative: 20  
 Best Local Similarity: 58.78% Mismatches: 38  
 Query Match: 54.79% Indels: 3  
 DB: 19 Gaps: 1

US-09-856-221-1 (1-449) x AAW56572 (1-2516)

QY 13 CACCTTATTGGCGGATAACCTTTATTTTCATTGGATAACGATTGGTCAGAACCCCGTTTA 72  
 |||||:||||| ||||| |||||:|||||  
 Db 1888 HisLeuLeuGlyAspLysProTyrLeuProLeuSerThrThrTrpSerAspProArgLeu 1907

QY 73 GAAGAAGCGCCAGTCACCAACCATTCGTGATCATTTATCAGCATAAATGCGCAACTGCGT 132  
 :: ||||| ||| :|||: ||| |||||  
 Db 1908 AspArgAlaAlaAspIleThrThrGlnAsnAlaHisAspSerAlaIleValAlaLeuArg 1927

QY 133 CAACCGCGCGCTTGGCGCGGAAA-----CGTACTGCAATTCGTTAACCGCTTTG 183  
 ||||| ||||| |||||:||||| |||||:|||||  
 Db 1928 GlnAsnIleProThrProAlaProLeuSerLeuArgSerAlaAsnThrLeuThrAspLeu 1947

QY 184 TTCCTTCCTCAGATAACAAAAAAGTTCAGTGGCAGAGCTTAGCACAGCCCTA 243  
 ||||| ||||| |||||:||||| |||||:|||||  
 Db 1948 PheLeuProGlnIleAsnGluValMetMetAsnTyrTrpGlnThrLeuAlaGlnArgVal 1967

QY 244 TATAACTTACGTCATAATCTGACAATTGATGTCAGCGTTGTCTATCCCATCTATGCG 303  
 ||||| ||||| |||||:||||| |||||:|||||  
 Db 1968 TyrAsnLeuArgHisAsnLeuSerIleAspGlyGlnProLeuTyrLeuProIleTyrAla 1987

QY 304 ACACGACAGATCCGTCGCTAGCTTACTGCTGCTGCGGTACCGCCCTCACAGCGGAGGG 363  
 ||||| ||||| |||||:||||| |||||:|||||  
 Db 1988 ThrProAlaAspProLysAlaLeuLeuSerAlaAlaValAlaThrSerGlnGlyGly 2007

QY 364 GATTTTCCTCGACAGTAATGCGCATGTACCGTTTTCGATATTCTGGAATAATGCCAAG 423  
 ||||| ||||| |||||:||||| |||||:|||||  
 Db 2008 LysLeuProGlnUserPheMetSerLeuTrpArgPheProHisMetLeuGluAsnAlaArg 2027

QY 424 TGGGAGTGACCCCACTGATACAG 447

Db 2028 GlyMetValSerGlnLeuThrGln 2035  
 |||||:||||| |||||  
 RESULT 3  
 AAB72609  
 ID AAB72609 standard; Protein; 2516 AA.  
 XX AC AAB72609;  
 XX 04-MAY-2001 (first entry)  
 XX Photorhabdus tcdA toxin.  
 XX TcdA; TcdA; insect toxin; plant; insect resistance.  
 XX Photorhabdus sp.  
 XX WO200111029-A1.  
 XX 15-FEB-2001.  
 XX 11-AUG-2000; 2000WO-US22237.  
 XX 11-AUG-1999; 99US-0148356.  
 XX (DOWC ) DOW AGROSCIENCES LLC.  
 XX Petell JK, Merlo DJ, Herman RA, Roberts JL, Guo L, Schafer BW;  
 PI Sukhapinda K, Merlo AO;  
 XX WPI: 2001-191536/19.  
 DR N-PSDB; AAF58778.  
 XX Novel polynucleotide sequence encoding insect toxins, useful for  
 PT producing transgenic plants having resistance to insects, especially  
 PT corn rootworm -  
 XX Disclosure; Page 51-61; 106pp; English.  
 XX The present invention provides the protein and coding sequences of  
 CC modified versions of the Photorhabdus tcdA and tcdA toxins. These are  
 CC suitable for expression in plants. The toxins are effective against  
 CC insects upon ingestion, and the sequences provided can be used to produce  
 CC transgenic plants with insect resistance. The present sequence is the  
 CC Photorhabdus tcdA protein.  
 XX Sequence 2516 AA;

Alignment Scores:  
 Pred. No.: 9,52e-46 Length: 2516  
 Score: 434.50 Matches: 87  
 Percent Similarity: 72.30% Conservative: 20  
 Best Local Similarity: 58.78% Mismatches: 38  
 Query Match: 54.79% Indels: 3  
 DB: 22 Gaps: 1

US-09-856-221-1 (1-449) x AAB72609 (1-2516)

QY 13 CACCTTATTGGCGGATAACCTTTATTTTCATTGGATAACGATTGGTCAGAACCCCGTTTA 72  
 |||||:||||| ||||| |||||:|||||  
 Db 1888 HisLeuLeuGlyAspLysProTyrLeuProLeuSerThrThrTrpSerAspProArgLeu 1907

QY 73 GAAGAAGCGCCAGTCACCAACCATTCGTGATCATTTATCAGCATAAATGCGCAACTGCGT 132  
 :: ||||| ||| :|||: ||| |||||  
 Db 1908 AspArgAlaAlaAspIleThrThrGlnAsnAlaHisAspSerAlaIleValAlaLeuArg 1927

QY 133 CAACCGCGCGCTTGGCGCGGAAA-----CGTACTGCAATTCGTTAACCGCTTTG 183  
 ||||| ||||| |||||:||||| |||||:|||||  
 Db 1928 GlnAsnIleProThrProAlaProLeuSerLeuArgSerAlaAsnThrLeuThrAspLeu 1947

QY 184 TTCCTTCCTCAGATAACAAAAAAGTTCAGTGGCAGAGCTTAGCACAGCCCTA 243  
 ||||| ||||| |||||:||||| |||||:|||||  
 Db 2008 LysLeuProGlnUserPheMetSerLeuTrpArgPheProHisMetLeuGluAsnAlaArg 2027

QY 424 TGGGAGTGACCCCACTGATACAG 447

QY 244 TATACTTACGTCATATCTGCAATGATGTCAGCGGTGTCTATTACCATCTATGCG 303  
 Db 1968 TyrAsnLeuArgHisAsnLeuSerIleAspGlyGlnProLeuTyrLeuProIleTyrAla 1987  
 QY 304 ACACGACGAGATCCGTCCTGCTACTGCTTACTGCTGCTGCTACCGCTCACAAGCGGAGGG 363  
 Db 1988 ThrProAlaAspProLysAlaLeuLeuSerAlaAlaValAlaThrSerGlnGlyGly 2007  
 QY 364 GATTGGCTCCGACGAGTAATGCGGATGATCCGTTTCCGATTATCTGGAATGCGCAAG 423  
 Db 2008 LysLeuProGluSerPheMetSerLeuTyrPargPheProHisMetLeuGluAsnAlaArg 2027  
 QY 424 TGGGAGTGACCAACTGATACAG 447  
 Db 2028 GlyMetValSerGlnLeuThrGln 2035

## RESULT 4

ID AAB72611 standard; Protein; 2517 AA.

XX AC AAB72611;

XX DT 04-MAY-2001 (first entry)

XX DE Modified Photorhabdus tcda toxin.

XX KW TcdA; TcBa; insect toxin; plant; insect resistance.

XX OS Photorhabdus sp.

XX OS Synthetic.

XX PN WO200111029-A1.

XX PD 15-FEB-2001.

XX PF 11-AUG-2000; 2000WO-US22237.

XX PR 11-AUG-1999; 99US-0148356.

XX PA (DOWC ) DOW AGROSCIENCES LLC.

XX PI Petell JK, Merlo DJ, Herman RA, Roberts JL, Guo L, Schafer BW;

XX PI Sukhapinda K, Merlo AO;

XX DR WPI; 2001-191536/19.

XX DR N-PSDB; AAF58780.

XX PT Novel polynucleotide sequence encoding insect toxins, useful for  
 PT producing transgenic plants having resistance to insects, especially  
 PT corn rootworm -

XX PS Claim 1; Page 72-83; 106pp; English.

XX CC The present invention provides the protein and coding sequences of  
 CC modified versions of the Photorhabdus tcda and tcba toxins. These are  
 CC suitable for expression in plants. The toxins are effective against  
 CC insects upon ingestion, and the sequences provided can be used to produce  
 CC transgenic plants with insect resistance. The present sequence is the  
 CC modified Photorhabdus tcda protein.

XX SQ Sequence 2517 AA;

## Alignment Scores:

Pred. No.:	9.52e-46	Length:	2517
Score:	434.50	Matches:	87
Percent Similarity:	72.30%	Conservative:	20
Best Local Similarity:	58.78%	Mismatches:	38
Query Match:	54.79%	Indels:	3
DB:	22	Gaps:	1

US-09-856-221-1 (1-449) x AAB72611 (1-2517)

QY 13 CACCTTATGGCGGATAACCTTATTTTCAITGGATAACGATTGGTCAGAACCCGTTTA 72  
 Db 1899 HisLeuLeuGlyAspLysProTyrLeuProLeuSerThrThrTrpSerAspProArgLeu 1908  
 QY 73 GAAGAAGCCGCCAGTCAAACCATTCGTGATCATATCAGCATAAATGCGGCACTGCGT 132  
 Db 1909 AspArgAlaAlaAspIleThrThrGlnAsnAlaHisAspSerAlaIleValAlaLeuArg 1928  
 QY 133 CAACGCGCGGCTTGGCGCGGAAA-----CGTACTGCAATTCGTTAAACCGCTTGG 183  
 Db 1929 GlnAsnIleProThrProAlaProLeuSerLeuArgSerAlaAsnThrLeuThrAspLeu 1948  
 QY 184 TTCCTTCTCAGATAAACAATAACTGCAAAATGCTACTGGCAGACGTTAGCACAAGCCTA 243  
 Db 1949 PheLeuProGlnIleAsnGluValMetAsnTyrTrpGlnThrLeuAlaGlnArgVal 1968  
 QY 244 TATACTTACGTCATATCTGACAAATGATGCTGACGCGTTTTCGATTATCTGGAATGCGCAAG 423  
 Db 2009 LysLeuProGluSerPheMetSerLeuTyrPargPheProHisMetLeuGluAsnAlaArg 2028  
 QY 424 TGGGAGTGACCAACTGATACAG 447  
 Db 2029 GlyMetValSerGlnLeuThrGln 2036

## RESULT 5

ID AAB72614

XX ID AAB72614 standard; Protein; 2537 AA.

XX AC AAB72614;

XX DT 04-MAY-2001 (first entry)

XX DE TcdA toxin-zein ER signal peptide fusion protein.

XX KW TcdA; TcBa; insect toxin; plant; insect resistance.

XX OS Chimeric - Photorhabdus sp.

XX OS Chimeric - Zea mays.

XX PN WO200111029-A1.

XX PD 15-FEB-2001.

XX PF 11-AUG-2000; 2000WO-US22237.

XX PR 11-AUG-1999; 99US-0148356.

XX PA (DOWC ) DOW AGROSCIENCES LLC.

XX PI Petell JK, Merlo DJ, Herman RA, Roberts JL, Guo L, Schafer BW;

XX PI Sukhapinda K, Merlo AO;

XX DR WPI; 2001-191536/19.

XX DR N-PSDB; AAF58783.

XX PT Novel polynucleotide sequence encoding insect toxins, useful for  
 PT producing transgenic plants having resistance to insects, especially  
 PT corn rootworm -

XX PS Example 1; Page 94-104; 106pp; English.

XX CC The present invention provides the protein and coding sequences of  
 CC modified versions of the Photorhabdus tcda and tcba toxins. These are  
 CC suitable for expression in plants. The toxins are effective against  
 CC insects upon ingestion, and the sequences provided can be used to produce



QY 430 GTGACCCAACTGATACAG 447  
 |||:|||||  
 Db 2036 ValSerGlnLeuThrGln 2041  
 RESULT 7  
 ID AAY97694 standard; Protein; 2376 AA.  
 XX AAY97694;  
 AC AAY97694;  
 XX  
 DT 19-JUN-2001 (first entry)  
 XX  
 DE SepA protein encoded by Serratia insecticidal protein complex gene.  
 XX  
 KW Insecticidal protein complex; amber disease; insect; Coleoptera;  
 KW pesticide; SepA protein.  
 XX  
 OS Serratia sp.  
 PN WO200116305-A2.  
 XX  
 PD 08-MAR-2001.  
 XX  
 PF 04-SEP-2000; 2000WO-NZ00174.  
 XX  
 PR 02-SEP-1999; 99NZ-0337610.  
 XX  
 PA (NZPA-) NEW ZEALAND PASTORAL AGRIC RES INST LTD.  
 XX  
 PI Glare TR, Hurst MRH, Jackson TA;  
 XX  
 DR WPI; 2001-169009/17.  
 DR N-PSDB; AAA91292.  
 XX  
 PT New nucleic acid encoding a polypeptide useful as a pesticide  
 PT especially for Coleoptera -  
 XX  
 PS Claim 24; Page 92-99; 109pp; English.  
 XX  
 CC This sequence represents the SepA protein encoded by the Serratia  
 CC insecticidal protein complex gene of the invention. The invention relates  
 CC to a gene encoding an insecticidal protein complex or a functional  
 CC fragment, a neutral mutation, or a homologue of the complex. The  
 CC polypeptides and nucleotides of the invention are used to induce amber  
 CC diseases or like conditions in insects, especially to those from the  
 CC order comprising Coleoptera, useful as a pesticide.  
 XX  
 SQ Sequence 2376 AA;

Alignment Scores:  
 Pred. No.: 1,81e-43 Length: 2376  
 Score: 416.50 Matches: 88  
 Percent Similarity: 72.97% Conservative: 20  
 Best Local Similarity: 59.46% Mismatches: 39  
 Query Match: 52.52% Indels: 2  
 DB: 22 Gaps: 1

US-09-856-221-1 (1-449) x AAY97694 (1-2376)

QY 2 GTGACGCACACCTATTGCGCGATAA-CCTATTTCATTGATACGATGGTCA 60  
 |||:|||||  
 Db 1756 ValGlnAlaLeuAsnLeuGlyAspGluProTyrIleSerPheAspAlaAspTrpSer 1775  
 |||:|||||  
 QY 61 GAACCCCGTTTAGAAGACCGCCAGTCAACACCATTCGTGATCATTTATCAGCATAAAATG 120  
 |||:|||||  
 Db 1776 AlaLeuThrLeuGlyAspAlaAlaSerGluValThrArgAspTyrGlnGluAlaLeu 1795  
 |||:|||||  
 QY 121 CGCAACCTGCGTCAACGCGCGCTTTCGCGGAAACGCTACTGCAAAATTCGTTAACCGCT 180  
 |||:|||||  
 Db 1796 LeuAlaValArgArgLeuValProAlaProGluThrArgThrAlaAsnSerLeuThrAla 1815  
 |||:|||||  
 QY 181 TTGTCCTCCTCAGATAAACAACAAAGTCTACTGGCAGACCTAGACACACGC 240  
 |||:|||||

Db 1816 LeuPheLeuProGlnGlnAsnGluValLeuLysGlyTyrTrpGlnThrIleuAlaGlnArg 1835  
 QY 241 CTATATAACTTACGTCATTAATCTGACAAATTGATGTCAGCGCTTGTTCATTACCATCTAT 300  
 |||:|||||  
 Db 1836 LeuHisAsnLeuArgHisAsnLeuSerIleAspGlyGlnProLeuSerLeuSerValTyr 1855  
 |||:|||||  
 QY 301 CGCACACGACAGATCCGCTACTGCTTACTAGTGTGCTCCGCTACCGCTCACAAGCGCGA 360  
 |||:|||||  
 Db 1856 AlaThrProSerGluProSerAlaLeuGlnSerAlaValValAsnSerAlaGlnGlyAla 1875  
 |||:|||||  
 QY 361 GGGATTTCCTCGGACAGTAATCCGATGACGTTTCCGATTATTCGGAAATGCC 420  
 |||:|||||  
 Db 1876 AlaAlaLeuProAlaAlaValMetProLeuTyrSerPheProValMetLeuGluAsnAla 1895  
 |||:|||||  
 QY 421 AAGTGGGAGTGCACCAACTGATA 444  
 |||:|||||  
 Db 1896 Arg---GlyMetValSerLeuLeu 1902  
 |||:|||||  
 RESULT 8  
 AAW56557  
 ID AAW56557 standard; Protein; 2504 AA.  
 XX  
 AC AAW56557;  
 XX  
 DT 07-AUG-1998 (first entry)  
 XX  
 DE Toxin Tcba, encoded by the tcbA gene from genomic region tcb.  
 XX  
 KW Photorhabdus luminescens W-14; nematode; symbiotic;  
 KW Heterorhabditis; tcb; tcb; tcc; tcd; insecticidal activity; toxin;  
 KW Lepidoptera; Coleoptera; Hymenoptera; Diptera; Dictyoptera; Acarina;  
 KW Homoptera; Southern; Western corn rootworm; Colorado potato beetle;  
 KW mealworm; boll weevil; turf grub; beetle armyworm; black cutworm;  
 KW cabbage looper; codling moth; corn earworm; European corn borer;  
 KW Tobacco hornworm; budworm.  
 XX  
 OS Photorhabdus luminescens.  
 XX  
 PN WO9808932-A1.  
 XX  
 PD 05-MAR-1998.  
 XX  
 PF 05-MAY-1997; 97WO-US07657.  
 XX  
 PR 06-NOV-1996; 96WO-US18003.  
 PR 28-AUG-1996; 96US-0705484.  
 PR 06-NOV-1996; 96US-0743699.  
 XX  
 PA (DOWC) DOWELANCO.  
 PA (WISC) WISCONSIN ALUMNI RES FOUND.  
 XX  
 PI Blackburn MB, Bowen DJ, Cliche TA, Ensign JC, Fatig R;  
 PI Ffrench-Constant RH, Guo L, Hey TD, Merlo DJ, Orr GL;  
 PI Petell J, Roberts JL, Rocheteau TA, Schoonover S;  
 PI Strickland JA, Sukhapinda K;  
 XX  
 DR WPI; 1998-179427/16.  
 DR N-PSDB; AAV29985.  
 XX  
 PT Isolated toxins from Photorhabdus luminescens strains - useful for  
 PT control of insect pests  
 XX  
 PS Claim 34; Pages 163-169; 321pp; English.  
 XX  
 CC The present sequence represents a protein named Tcba of the bacterium  
 CC Photorhabdus luminescens (W-14). This is a symbiotic bacterium of the  
 CC nematodes of the Heterorhabditis genus. The bacterium has at least 4  
 CC distinct genomic regions, tcb, tcb, tcc, and tcd. Peptide products are  
 CC produced from these regions that are associated with insecticidal  
 CC activity. The native toxins are secreted proteins. The proteins are  
 CC toxic to insects upon exposure and especially when ingested. The  
 CC nucleic acid sequence can be used to produce transgenic plants,  
 CC baculoviruses or microbial hosts for toxin production. They can be used



CC to control insects pests from the Lepidoptera, Coleoptera, Hymenoptera,  
CC Diptera, Dictyoptera, Acarina or Homoptera orders, especially the  
CC Southern or Western corn rootworm, Colorado potato beetle, mealworm,  
CC boll weevil, turf grub, beetle armyworm, black cutworm, cabbage looper,  
CC codling moth, corn earworm, European corn borer or tobacco hornworm  
CC or budworm.

XX  
S0 Sequence 2504 AA:

Alignment Scores:		
Pred. No.:	2.87e-43	2504
Score:	415.00	90
Percent Similarity:	71.33%	Matches: 17
Best Local Similarity:	60.00%	Mismatches: 41
Query Match:	52.33%	Indels: 3
DB:	19	Gaps: 1

US-09-856-221-1 (1-449) x AAW56557 (1-2504)

QY	5	CAGGCACTCACCTTTATTGGCGGATAA-CCTTATTTTTCATTGGATAACGATTGGTCAGAA	63
Db	1880	GlnAlaLeuAsnLeuLeuGlyAspGluProGlnValMetLeuSerThrTrpAlaAsn	1899
QY	64	CCCGTWTAGAAAGCCGCCAGTCAAAACCATTCGGTGATCATATCACCATAAATCGG	123
Db	1900	ProThrLeuGlyAsnAlaAlaSerIysThrThrGlnGlnValArgGlnValLeuThr	1919
QY	124	CAACTGGGT-----CAAGCGCGCGCTTCCGCGGAAAGCTACTGCAAAATTCGTTAAAC	177
Db	1920	GlnLeuArgLeuAsnSerArgValLysThrProLeuLeuGlyThrAlaAsnSerLeuThr	1939
QY	178	GCTTTGTTCCTCTCAGATAAACAACAAACTGCAAAAGTTACTGGCAGACGTTAGCAGAA	237
Db	1940	AlaLeuPheLeuProGlnGluAsnSerLysLeuLysGlyTrpArgThrLeuAlaGln	1959
QY	238	CGCTATATAACTTACGTCATTAATCTGACAATTCATGGTCAGCGGTTCATTACCCATC	297
Db	1960	ArgMetPheAsnLeuArgHisAsnLeuSerIleAspGlyGlnProLeuSerLeuProLeu	1979
QY	298	TATGGACACACGACATCCGTCCTACTGTTTAGTCTGCTCCGTCACGCCCTCACAAGGC	357
Db	1980	TyrAlaLysProAlaAspProLysAlaLeuLeuSerAlaAlaValSerAlaSerGlnGly	1999
QY	358	GGAGGGATTTGGCTCGGACAGTACGTCCGAGTACCGTTTTTCCGATTATTTCTGGAAAT	417
Db	2000	GlyAlaAspLeuProLysAlaProLeuThrIleHisArgPheProGlnMetLeuGlnGly	2019
QY	418	GCCAAAGTGGGAGTGCACCAACTGATACAG	447
Db	2020	AlaArgGlyLeuValAsnGlnLeuIleGln	2029

RESULT 9  
AAB72610

AAB/2010  
ID AAB72610

XX  
0107/BBY  
01

AC AAB72610;

XX

DT 04-MAY-2011

XX  
25-4-1-1-1

DE Photorhabdovirus

XX  
KW  
Toda. TchXX  
TCQA; TCD

OS Photorhabd

XX  
XX  
XXXXXXX

PN WO2001110

XX

PD 15-FEB-20

XX DE 11-AUG-30

PF  
II-AUG-20  
YY

11-AUG-19

XX  
II-806-TT

11

PA	(DOWC ) DOW AGROSCIENCES LLC.	
XX		
XX	Petell JK, Merlo DJ, Herman RA, Roberts JL, Guo L, Schafer BW;	
PI	Sukhapinda K, Merlo AO;	
XX		
XX	WPI; 2001-191536/19.	
DR	N-PSDB; AAF58779.	
DR		
DR		
XX		
XX	Novel polynucleotide sequence encoding insect toxins, useful for	
PT	producing transgenic plants having resistance to insects, especially	
PT	corn rootworm	
XX		
XX	Disclosure; Page 62-72; 106pp; English.	
XX		
CC	The present invention provides the protein and coding sequences of	
CC	modified versions of the photorhabdus Tcda and TcBa toxins. These are	
CC	suitable for expression in plants. The toxins are effective against	
CC	insects upon ingestion, and the sequences provided can be used to produce	
CC	transgenic plants with insect resistance. The present sequence is the	
CC	photorhabdus TcBa protein.	

XX	SQ	Sequence	2504	AA
Alignment Scores:				
Pred. No.:		2.87e-43	Length:	2504
Score:		415.00	Matches:	90
Percent Similarity:		71.33%	Conservative:	17
Best Local Similarity:		60.00%	Mismatches:	41
Query Match:		52.33%	Indels:	3
DB:		22	Gaps:	1

US-09-856-221-1 (1-449) x AAB72610 (1-2504)

[illegible]

RESULT 10  
AAB72612

AAB72612  
ID AAB72612

XX  
XX

AC AAB72612;

XX

DT 04-MAY-20

XX

DE Modified Photorhabdus tcba toxin.  
 XX Tcda; Tcda; insect toxin; plant; insect resistance.  
 KW Photorhabdus sp.  
 OS Synthetic.  
 XX WO200111029-A1.  
 XX 15-FEB-2001.  
 XX 11-AUG-2000; 2000WO-US22237.  
 XX 11-AUG-1999; 99US-0148356.  
 XX (DOWC ) DOW AGROSCIENCES LLC.  
 XX Petell JK, Merlo DJ, Herman RA, Roberts JL, Guo L, Schafer BW;  
 PI Sukhapinda K, Merlo AO;  
 XX WPI; 2001-191536/19.  
 DR N-PSDB; AAF58781.  
 XX Novel polynucleotide sequence encoding insect toxins, useful for  
 PT producing transgenic plants having resistance to insects, especially  
 PT corn rootworm -  
 XX Claim 1; Page 83-93; 106pp; English.  
 XX The present invention provides the protein and coding sequences of  
 CC modified versions of the Photorhabdus Tcda and Tcda toxins. These are  
 CC suitable for expression in plants. The toxins are effective against  
 CC insects upon ingestion, and the sequences provided can be used to produce  
 CC transgenic plants with insect resistance. The present sequence is the  
 CC modified Photorhabdus Tcda protein.  
 XX Sequence 2505 AA;  
 SQ  
 Alignment Scores:  
 Pred. No.: 6,92e-43 Length: 2505  
 Score: 412.00 Matches: 90  
 Percent Similarity: 71.33% Conservative: 17  
 Best Local Similarity: 60.00% Mismatches: 41  
 Query Match: 51.95% Indels: 3  
 DB: 22 Gaps: 1  
 US-09-856-221-1 (1-449) x AAB72612 (1-2505)  
 QY 5 CAGGCACCTTATTTGGGCGATAA-CCTTATTTTTCATTGGATACGATTGGTCAGAA 63  
 Db 1881 GlnAlaLeuAsnLeuLeuGlyAspGluProGlnValMetLeuSerThrTrpAlaAsn 1900  
 QY 64 CCGCGTTTAAAGAGCGCCAGTCAACCATTCGTGATCATATATCAAGATAAATGCGG 123  
 Db 1901 ProThrLeuGlyAsnAlaAlaSerLysThrThrGlnGlnValArgGlnGlnValLeuThr 1920  
 QY 124 CAACGCGGT-----CAACGCGCGCGCTTGGCGGCGAAACGCTACTGCAAAATTCGTTAAC 177  
 Db 1921 GlnLeuArgLeuAsnSerArgValLysThrProLeuLeuGlyThrAlaAsnSerLeuThr 1940  
 QY 178 GCTTTGTTCTCTCTCAGATAAATAAAGTCAAGTTACTGCGACGTTTATGACACAA 237  
 Db 1941 AlaLeuPheLeuProGlnGlnValSerLysLeuLysGlyTyrTrpArgThrLeuAlaGln 1960  
 QY 238 CGCTATATACCTACGTCATAATCTGACAAATGATGTCAGCGGTTGTCATTCACCATC 297  
 Db 1961 ArgMetPheAsnLeuArgHisAsnLeuSerIleAspGlyGlnProLeuSerLeuProLeu 1980  
 QY 298 TATGCGACACGACGAGTCCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 357  
 Db 1981 TyrAlaLysProAlaAspProLysAlaLeuLeuSerAlaAlaValSerAlaSerGlnGly 2000  
 QY 358 GGAGGGGATTTCCTCGACAGTAATAGCCGCGATGACCGGTTTTCGATTATTCGGAAT 417

Db 2001 GlyAlaAspLeuProLysAlaProLeuThrIleHisArgPheProGlnMetLeuGluGly 2020  
 QY 418 GCCAAGTGGGAGTGACCCCACTGATACAG 447  
 Db 2021 AlaArgGlyLeuValAsnGlnLeuIleGln 2030  
 RESULT 11  
 AAW17871  
 ID AAW17871 standard; Protein; 2504 AA.  
 XX  
 AC AAW17871;  
 XX  
 DT 29-JAN-1998 (first entry)  
 XX  
 DE Photorhabdus luminescens insect toxin protein Tcda.  
 XX  
 KW Insecticide; insect; toxin; pest control; biological control;  
 KW Photorhabdus luminescens; Tcda; Southern corn rootworm;  
 KW Colorado potato beetle; Western corn rootworm; meal worm;  
 KW boll weevil; turf grub; Coleoptera; beet armyworm; black cutworm;  
 KW cabbage looper; codling moth; corn earworm; European corn borer;  
 KW tobacco hornworm; tobacco budworm; Lepidoptera; Hymenoptera;  
 KW Diptera, Dictyoptera; Acarina; Homoptera.  
 XX  
 OS Photorhabdus luminescens strain W-14 (ATCC 55397).  
 XX  
 PN WO9717432-A1.  
 XX  
 PD 15-MAY-1997.  
 XX  
 PF 06-NOV-1996; 96WO-US18003.  
 XX  
 PR 28-AUG-1996; 96US-0705484.  
 PR 06-NOV-1995; 95US-0007255.  
 PR 28-FEB-1996; 96US-0608423.  
 XX  
 PA (WISC ) WISCONSIN ALUMNI RES FOUND.  
 XX  
 PI Blackburn MB, Bowen DJ, Ciche TA, Ensign JC, Fatig R;  
 PI Ffrench-Constant RH, Guo L, Hey TD, Merlo DJ, Orr GL;  
 PI Petell J, Roberts JL, Rocheleau TA, Schoonover S;  
 PI Strickland JA;  
 XX  
 DR WPI; 1997-281022/25.  
 DR N-PSDB; AAT68836.  
 XX  
 PT Photorhabdus sp. insecticidal protein toxins and DNA encoding them -  
 PT can be genetically engineered into insect larvae food and plants for  
 PT insect control  
 XX  
 PS Claim 8; Page 119-129; 276pp; English.  
 XX  
 CC This polypeptide comprises a specifically claimed 280.6 kDa  
 CC insecticidal toxin, Tcda, of Photorhabdus luminescens that is  
 CC proteolytically cleaved into 207.6 kDa (see AAW18302) and 62.9 kDa  
 CC polypeptides (see AAW18303). The Tcda polypeptide can be expressed  
 CC in host cells using a gene (see AAT68836) isolated from a genomic  
 CC library. Claimed toxin proteins of P. luminescens (see AAW17871,  
 CC AAW17884-89, AAW17899-900, AAW18301-06) can be applied to, or  
 CC genetically engineered into, insect larvae food and plants for insect  
 CC control. The Photorhabdus toxins are particularly effective against  
 CC Southern corn rootworm, Colorado potato beetle, Western corn rootworm,  
 CC meal worm, boll weevil and turf grub (Coleoptera), beet armyworm, black  
 CC cutworm, cabbage looper, codling moth, corn earworm, European corn  
 CC borer, tobacco hornworm and tobacco budworm (Lepidoptera), and  
 CC are also active against insects of the orders Hymenoptera, Diptera  
 CC Dictyoptera, Acarina and Homoptera. (All claimed).  
 XX  
 SQ Sequence 2504 AA;  
 Alignment Scores:  
 Pred. No.: 1.24e-42 Length: 2504



KW Heterorhabditis; tea; tcb; tcc; tcd; insecticidal activity; toxin; Lepidoptera; Coleoptera; Hymenoptera; Diptera; Dictyoptera; Acarina; Homoptera; Southern; Western corn rootworm; Colorado potato beetle; mealworm; boll weevil; turf grub; beetle armyworm; black cutworm; cabbage looper; codling moth; corn earworm; European corn borer; Tobacco hornworm; budworm.

XX  
OS Photorhabdus luminescens.

XX  
PN WO9808932-A1.

XX  
PD 05-MAR-1998.

XX  
PF 05-MAY-1997; 97WO-US07657.

XX  
PR 06-NOV-1996; 96WO-US18003.

PR 28-AUG-1996; 96US-0705484.

PR 06-NOV-1996; 96US-0743699.

XX  
PA (DOWC ) DOWELANCO.

PA (WISC ) WISCONSIN ALUMNI RES FOUND.

XX  
PI Blackburn MB, Bowen DJ, Cliche TA, Ensign JC, Fatig R; French-Constant RH, Guo L, Hey TD, Merlo DJ, Orr GL; Petell J, Roberts JL, Rocheleau TA, Schoonover S; Strickland JA, Sukhapinda K;

XX  
DR WPI; 1998-179427/16.

DR N-PSDB; AAV29931.

XX  
PT Isolated toxins from Photorhabdus luminescens strains - useful for control of insect pests

XX  
PS Claim 34; Pages 245-247; 321pp; English.

XX  
CC The present sequence represents a protein named TcdAIII of the bacterium Photorhabdus luminescens (W-14). This is a symbiotic bacterium of the nematodes of the Heterorhabditis genus. The bacterium has at least 4 distinct genomic regions, tea, tcb, tcc, and tcd. Peptide products are produced from these regions that are associated with insecticidal activity. The native toxins are secreted proteins. The proteins are toxic to insects upon exposure and especially when ingested. The nucleic acid sequence can be used to produce transgenic plants, baculoviruses or microbial hosts for toxin production. They can be used to control insects pests from the Lepidoptera, Coleoptera, Hymenoptera, Diptera, Dictyoptera, Acarina or Homoptera orders, especially the Southern or Western corn rootworm, Colorado potato beetle, mealworm, boll weevil, turf grub, beetle armyworm, black cutworm, cabbage looper, codling moth, corn earworm, European corn borer or tobacco hornworm or budworm.

XX  
SQ Sequence 579 AA;

Alignment Scores:  
Pred. No.: 9.32e-36 Length: 579  
Score: 354.00 Matches: 68  
Percent Similarity: 83.51% Conservativity: 13  
Best Local Similarity: 70.10% Mismatches: 16  
Query Match: 44.64% Indels: 0  
DB: 19 Gaps: 0

US-09-856-221-1 (1-449) x AAW56574 (1-579)

QY 157 CGTACTGCAATTCGTTAACCGTTTGTCCTCTCAGATAAACAACAACTGCAAGT 216  
|||||  
D 2 ArgSerAlaAsnThrLeuThrAspLeuProGlnIleAsnGluValMetMetAsn 21  
|||||  
QY 217 TACTGCGACGTTAGCAACACGCTATATACCTTACGTCATATCTGACAATGTAGT 276  
|||||  
D 22 TyrTrpGlnThrLeuAlaGlnArgValTyrAsnLeuArgHisAsnLeuSerIleAspGly 41  
|||||  
QY 277 CAGCGGTGTCATTACCCATCTATGCGACACCGACAGATCGCTAGCTAGTGT 336  
|||||

Db 42 GlnProLeuTyrLeuProLeuTyrAlaThrProAlaaspProLysAlaLeuLeuSerAla 61

QY 337 CCCCTCACCCCTCACAGCGGAGGATTGCTCGGACAGTAATGCGATGTACCGT 396  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 62 AlaValAlaThrSerGlnGlyGlyLysLeuProGluSerPheMetSerLeuTrpArg 81

QY 397 TTTCGATTATTCTGGAAATGCCAAGTGGGAGTGACCCAACTGATACAG 447  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 82 PheProHisMetLeuGluAsnAlaArgGlyMetValSerGlnLeuThrGln 98

RESULT 14

AAW18303

ID AAW18303 standard; Protein: 573 AA.

XX  
AC AAW18303;

XX  
DT 30-JAN-1998 (first entry)

XX  
DE Photorhabdus luminescens insect toxin TcdAIII.

XX  
KW Insecticide; insect; toxin; pest control; biological control; Photorhabdus luminescens; TcbA; Southern corn rootworm; Colorado potato beetle; Western corn rootworm; meal worm; boll weevil; turf grub; Coleoptera; beet armyworm; black cutworm; cabbage looper; codling moth; corn earworm; European corn borer; tobacco hornworm; tobacco budworm; Lepidoptera; Hymenoptera; Diptera, Dictyoptera; Acarina; Homoptera.

XX  
OS Photorhabdus luminescens strain W-14 (ATCC 55397).

XX  
PN WO9717432-A1.

XX  
PD 15-MAY-1997.

XX  
PF 06-NOV-1996; 96WO-US18003.

XX  
PR 28-AUG-1996; 96US-0705484.

PR 06-NOV-1995; 95US-0007255.

PR 28-FEB-1996; 96US-0608423.

XX  
PA (WISC ) WISCONSIN ALUMNI RES FOUND.

XX  
PI Blackburn MB, Bowen DJ, Cliche TA, Ensign JC, Fatig R; French-Constant RH, Guo L, Hey TD, Merlo DJ, Orr GL; Petell J, Roberts JL, Rocheleau TA, Schoonover S; Strickland JA;

XX  
DR WPI; 1997-281022/25.

DR N-PSDB; AAT68847.

XX  
PT Photorhabdus sp. insecticidal protein toxins and DNA encoding them - can be genetically engineered into insect larvae food and plants for insect control

PT  
PT  
XX  
PS Claim 34; Page 229-231; 276pp; English.

XX  
CC This polypeptide comprises the 62.9 kDa TcdAIII insecticidal toxin protein of Photorhabdus luminescens W-14. Its sequence was deduced from a genomic DNA clone (AAT68847). TcdAIII is a proteolytic cleavage product of TcbA (see AAW17871). Claimed toxins of P. luminescens (see AAW17871, AAW17884-89, AAW17899-900, AAW18301-06) can be produced by recombinant DNA methods and applied to, or genetically engineered into, insect larvae food and plants for insect control.

CC The toxins are particularly effective against Southern corn rootworm, Colorado potato beetle, Western corn rootworm, meal worm, boll weevil and turf grub (Coleoptera), beet armyworm, black cutworm, cabbage looper, codling moth, corn earworm, European corn borer, tobacco hornworm and tobacco budworm (Lepidoptera), and are also active against insects of the orders Hymenoptera, Diptera, Dictyoptera, Acarina and Homoptera. (All claimed).

XX  
SQ Sequence 573 AA;

Alignment Scores:

Pred. No.: 1.24e-35 Length: 573  
Score: 353.00 Matches: 68  
Percent Similarity: 84.38% Conservative: 13  
Best Local Similarity: 70.83% Mismatches: 15  
Query Match: 44.51% Indels: 0  
DB: 18 Gaps: 0

US-09-856-221-1 (1-449) x AAW18303 (1-573)

QY 160 ACTGCAATTCGTAACCGCTTTGCTCTCCTCAGATAAACAACAACTGCAAGTTAC 219  
Db 3 ThrAlaAsnSerLeuThrAlaLeuPheLeuProGlnGluAsnSerLysLeuGlyTyr 22  
QY 220 TGGCAGACGTTAGCACAAGCGCTATATACCTACGTCATATCTGACATTCATGCTGTCAG 279  
Db 23 TrpArgThrLeuAlaGlnArgMetPheAsnLeuArgHisAsnLeuSerIleAspGlyGln 42  
QY 280 CCGTGTGTCATACCATCTATCGGACACGACGATCCGTCGCTACTGCTTACTGCTGCC 339  
Db 43 ProLeuSerLeuProLeuTyrAlaLysProAlaAspProLysAlaLeuLeuSerAlaAla 62  
QY 340 GTACCGCGCTCACAAGCGGAGGATTTGCTCGGACAGTAATGCGGATGTACCGTTT 399  
Db 63 ValSerAlaSerGlnGlyAlaAspLeuProLysAlaProLeuThrIleHisArgPhe 82  
QY 400 CCGATTATTCTGGAATGCAAGTGGGAGTGACCACTGATACAG 447  
Db 83 ProGlnMetLeuGluGlyAlaArgGlyLeuValAsnGlnLeuIleGln 98

RESULT 15

AAW56559  
ID AAW56559 standard; Protein; 573 AA.

XX AC AAW56559;

XX DT 07-AUG-1998 (first entry)

XX DE Toxin TcbAII, encoded by the tcbA gene from genomic region tcb.

XX KW Photorhabdus luminescens W-14; nematode; symbiotic;  
XX KW Heterorhabdus; tcb; tcb; tcc; tcd; insecticidal activity; toxin;  
XX KW Lepidoptera; Coleoptera; Hymenoptera; Diptera; Dictyoptera; Acarina;  
XX KW Homoptera; Southern; Western corn rootworm; Colorado potato beetle;  
XX KW mealworm; boll weevil; turf grub; beetle armyworm; black cutworm;  
XX KW cabbage looper; codling moth; corn earworm; European corn borer;  
XX KW Tobacco hornworm; budworm.

XX OS Photorhabdus luminescens.

XX PN WO9808932-A1.

XX PD 05-MAR-1998.

XX PF 05-MAY-1997; 97WO-US07657.

XX PR 06-NOV-1996; 96WO-US18003.

XX PR 28-AUG-1996; 96US-0705484.

XX PR 06-NOV-1996; 96US-0743699.

XX PA (DOWC ) DOWELANCO.

XX PA (WISC ) WISCONSIN ALUMNI RES FOUND.

XX PI Blackburn MB, Bowen DJ, Cliche TA, Ensign JC, Fatig R;

XX PI Zfirench-Constant RH, Guo L, Hey TD, Merlo DJ, Orr GL;

XX PI Petell J, Roberts JL, Rocheleau TA, Schoonover S;

XX PI Strickland JA, Sukhapinda K;

XX DR WPI; 1998-179427/16.

XX DR N-PSDB; AAV2998.

XX Isolated toxins from Photorhabdus luminescens strains - useful for

XX control of insect pests

XX Claim 34; Pages 261-263; 321pp; English.

XX The present sequence represents a protein named TcbAII of the bacterium  
CC Photorhabdus luminescens (W-14). This is a symbiotic bacterium of the  
CC nematodes of the Heterorhabdus genus. The bacterium has at least 4  
CC distinct genomic regions, tcb, tcb, tcc, and tcd. Peptide products are  
CC produced from these regions that are associated with insecticidal  
CC activity. The native toxins are secreted proteins. The proteins are  
CC toxic to insects upon exposure and especially when ingested. The  
CC nucleic acid sequence can be used to produce transgenic plants,  
CC baculoviruses or microbial hosts for toxin production. They can be used  
CC to control insects pests from the Lepidoptera, Coleoptera, Hymenoptera,  
CC Diptera, Dictyoptera, Acarina or Homoptera orders, especially the  
CC Southern or Western corn rootworm, Colorado potato beetle, mealworm,  
CC boll weevil, turf grub, beetle armyworm, black cutworm, cabbage looper,  
CC codling moth, corn earworm, European corn borer or tobacco hornworm  
CC or budworm.

XX SO Sequence 573 AA:

Alignment Scores:

Pred. No.: 1.24e-35 Length: 573  
Score: 353.00 Matches: 68  
Percent Similarity: 84.38% Conservative: 13  
Best Local Similarity: 70.83% Mismatches: 15  
Query Match: 44.51% Indels: 0  
DB: 18 Gaps: 0

US-09-856-221-1 (1-449) x AAW56559 (1-573)

QY 160 ACTGCAATTCGTTAACCGCTTTGCTCTCCTCAGATAAACAACAACTGCAAGTTAC 219  
Db 3 ThrAlaAsnSerLeuThrAlaLeuPheLeuProGlnGluAsnSerLysLeuGlyTyr 22  
QY 220 TGGCAGACGTTAGCACAAGCGCTATATACCTACGTCATATCTGACATTCATGCTGTCAG 279  
Db 23 TrpArgThrLeuAlaGlnArgMetPheAsnLeuArgHisAsnLeuSerIleAspGlyGln 42  
QY 280 CCGTGTGTCATACCATCTATCGGACACGACGATCCGTCGCTACTGCTTACTGCTGCC 339  
Db 43 ProLeuSerLeuProLeuTyrAlaLysProAlaAspProLysAlaLeuLeuSerAlaAla 62  
QY 340 GTACCGCGCTCACAAGCGGAGGATTTGCTCGGACAGTAATGCGGATGTACCGTTT 399  
Db 63 ValSerAlaSerGlnGlyAlaAspLeuProLysAlaProLeuThrIleHisArgPhe 82  
QY 400 CCGATTATTCTGGAATGCAAGTGGGAGTGACCACTGATACAG 447  
Db 83 ProGlnMetLeuGluGlyAlaArgGlyLeuValAsnGlnLeuIleGln 98

Search completed: January 15, 2003, 15:50:23

Job time : 53.5673 secs



GenCore version 5.1.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: January 15, 2003, 15:45:23 ; Search time 14.4662 Seconds  
(without alignments)  
1826.456 Million cell updates/sec

Title: us-09-856-221-1

Perfect score: 793  
Sequence: 1 tgtcagcagcactcattat.....gtgacccaactgatacagtt 449

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+\_n2p\_model -DEV=xlp  
-Q=/cgn2\_1/USPTO.spool/US09856221/runat\_15012003\_153926\_12521/app\_query.fasta\_1.1948  
-DB=Issued\_Patents\_AA -QFMT=fastan -SUFFIX=rai -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptlo -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09856221.ecgn.1.1.24.ernat.15012003.153926.12521 -NCPD=6 -ICPD=3  
-NO\_XLPVX -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep.\*  
3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep.\*  
4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep.\*  
5: /cgn2\_6/ptodata/1/1aa/PTCUS.COMB.pep.\*  
6: /cgn2\_6/ptodata/1/1aa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	428.5	54.0	2522	4	US-09-251-645-13
2	71.5	9.0	942	3	US-09-074-579-1
3	71.5	9.0	942	4	US-09-388-774-1
4	71	9.0	2441	1	US-08-194-468-2
5	71	9.0	2441	3	US-08-961-739-2
6	71	9.0	2441	4	US-08-514-247A-8
7	70	8.8	216	2	US-08-840-683-9
8	70	8.8	216	2	US-08-555-722-9
9	70	8.8	216	4	US-09-384-301-9
10	70	8.8	983	4	US-09-134-001C-3814
11	69.5	8.8	342	3	US-08-978-456-2
12	69.5	8.8	342	4	US-09-369-700-2

13	69	8.7	201	1	US-08-444-083-8	Sequence 8, Appli
14	69	8.7	201	1	US-08-286-304-8	Sequence 8, Appli
15	69	8.7	201	1	US-08-442-745-8	Sequence 8, Appli
16	69	8.7	201	1	US-08-443-129-8	Sequence 8, Appli
17	69	8.7	201	1	US-08-443-952-8	Sequence 8, Appli
18	69	8.7	201	1	US-08-443-130-8	Sequence 8, Appli
19	69	8.7	201	1	US-08-792-019B-11	Sequence 11, Appli
20	69	8.7	201	3	US-09-106-182-4	Sequence 4, Appli
21	69	8.7	201	3	US-08-988-819-11	Sequence 11, Appli
22	69	8.7	201	3	US-08-898-911-8	Sequence 8, Appli
23	69	8.7	201	4	US-09-016-534-11	Sequence 11, Appli
24	69	8.7	201	5	PCT-US95-04467-8	Sequence 8, Appli
25	68	8.6	448	3	US-08-476-509B-2	Sequence 2, Appli
26	68	8.6	486	3	US-08-348-518C-2	Sequence 2, Appli
27	68	8.6	1274	4	US-09-095-443-2	Sequence 2, Appli
28	67.5	8.5	834	3	US-08-539-205A-6	Sequence 6, Appli
29	67.5	8.6	837	1	US-07-923-976-2	Sequence 2, Appli
30	66.5	8.5	951	1	US-08-162-809-2	Sequence 2, Appli
31	66	8.3	698	4	US-09-134-001C-3632	Sequence 3632, Ap
32	66	8.3	1349	3	US-08-938-291A-6	Sequence 6, Appli
33	65	8.3	299	4	US-09-370-976-2	Sequence 2, Appli
34	65	8.2	660	4	US-09-462-606-57	Sequence 57, Appli
35	65	8.2	1417	2	US-08-559-303B-78	Sequence 78, Appli
36	65	8.2	1417	3	US-08-781-891-78	Sequence 78, Appli
37	65	8.2	1417	4	US-09-175-828-78	Sequence 78, Appli
38	64.5	8.1	506	4	US-08-942-012B-26	Sequence 26, Appli
39	64.5	8.1	624	2	US-08-879-561-3	Sequence 3, Appli
40	64.5	8.1	659	1	US-08-240-049B-16	Sequence 16, Appli
41	64.5	8.1	659	1	US-08-259-148A-20	Sequence 20, Appli
42	64.5	8.1	659	2	US-08-484-054-20	Sequence 20, Appli
43	64.5	8.1	659	2	US-07-876-941A-20	Sequence 14, Appli
44	64.5	8.1	659	4	US-08-477-292-14	Sequence 20, Appli
45	64.5	8.1	659	4	US-07-870-985A-20	Sequence 20, Appli

## ALIGNMENTS

RESULT 1  
US-09-251-645-13  
; Sequence 13, Application US/09251645  
; Patent No. 6281413  
; GENERAL INFORMATION:  
; APPLICANT: Kramer, Vance C.  
; APPLICANT: Morgan, Michael K.  
; APPLICANT: Anderson, Arne R.  
; APPLICANT: Hart, Hope  
; APPLICANT: Warren, Gregory W.  
; APPLICANT: Dunn, Martha  
; APPLICANT: Chen, Jeng S.  
; TITLE OF INVENTION: NOVEL INSECTICIDAL TOXINS FROM PHOTORHABDUS LUMINESCENS  
; FILE REFERENCE: CGC1963/A  
; CURRENT APPLICATION NUMBER: US/09/251.645  
; CURRENT FILING DATE: 1999-02-17  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 13  
; LENGTH: 2522  
; TYPE: PRT  
; ORGANISM: Photorhabdus luminescens  
US-09-251-645-13

Alignment Scores:  
Pred. No.: 1.2e-46  
Score: 428.50  
Length: 2522  
Matches: 86  
Percent Similarity: 73.97%  
Conservative: 86  
Best Local Similarity: 58.90%  
Mismatches: 37  
Query Match: 54.04%  
Indels: 1  
Gaps: 1  
DB: 4

US-09-856-221-1 (1-449) x US-09-251-645-13 (1-2522)

QY 13 CACCTTATGGCGATAACCTTATTTTCATGTGATAACGATGGTCAGAACCCGGTTTA 72





```

/ sequence 4, Application US/08/0961739A
/ Patent No. 6063583
/ GENERAL INFORMATION:
/ APPLICANT: Montminy, Marc R.
/ TITLE OF INVENTION: Methods for Treating Diabetes Mellitus
/ FILE REFERENCE: SALK1650-1
/ CURRENT APPLICATION NUMBER: US/08/961.739A
/ CURRENT FILING DATE: 1997-10-31
/ EARLIER APPLICATION NUMBER: US 194,468
/ EARLIER FILING DATE: 1994-02-10
/ NUMBER OF SEQ ID NOS: 4
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 2
/ LENGTH: 2441

```

```

Percent Similarity: 38.97%      Conservative: 19
Best Local Similarity: 25.00%    Mismatches: 45
Query Match: 8.95%             Indels: 38
DB: 4                          Gaps: 7

US-09-856-221-1 (1-449) x US-09-514-247A-8 (1-2441)

Oy 76 GAAGCGCGAGTCAACCATTCGGATCATATTACGACATAAAATCGCGCRA----- 126
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 593 GluHisValThrGlnAspLeuArgSerHisLeuValHisLysLeuValGlnAlaIlePhe 612
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Oy 127 --CTGCGTCAACGGCGCGCTCCCGCGGAAACGACTGCAAAATTCGTTAACCGCTTG 183
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 613 ProThrProAspProAlaAlaLeuLysAspArgArgMetGluAsn----- 627
Oy 184 TTCCTTCCTCAGATAACCAAAAACTGCAA----- 213
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 628 ---LeuValAlaValAlaLysLysValGluGlyAspMetTyrGluSerAlaAsnSerArg 646
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Oy 214 ---AGTTACTGCGACAGCTTAGCAACGCGCTATAATACTTACGT----- 255
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 647 AspGluTyrTyrHisLeuLeuAlaGluLysIleLeuGlyAsnGlnProAlaLeuPro 666
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Oy 256 -----CATAAATCTGCACAAATGATGGTCACGCCGTTGTCAATTACCCATC 297
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 667 LysArgArgThrArgLeuHisLysGlnGlyIleLeuGlyAsnGlnProAlaLeuPro 685
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Oy 298 TATGCGACACGACAGATCCGTCGCTACTGCTTAGTGCCTGCGCTCACCGCTCACAGGC 357
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 686 ---AlaSerGlyAlaGlnProProVal---IleProAlaGlnSerValArgProPro 703
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Oy 358 GGAGGGGATTTCCTCGGACACGATGATGCGGATGACGCTTTTCCGATT 405
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 704 AsnGlyProLeuPro-----LeuProValAsnArgMetGlnVal 716
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

RESULT 7
US-08-840-683-9
; Sequence 9, Application US/08840683
; Patent No. 5821051
; GENERAL INFORMATION:
; APPLICANT: Androphy, Elliot J.
; APPLICANT: Chen, Jason J.
; TITLE OF INVENTION: E6-BINDING PROTEINS
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII (text)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/840,683
; FILING DATE: 29-APR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/273,059
; FILING DATE: 08-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Louis
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: NEP-003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 216 amino acids

```

```

; TYPE: amino acid
; MOLECULE TYPE: linear
; FRAGMENT TYPE: N-terminal
US-08-840-683-9
Alignment Scores:
Pred. No.: 1.51 Length: 216
Score: 70.00 Matches: 25
Percent Similarity: 43.59% Conservative: 9
Best Local Similarity: 32.05% Mismatches: 34
Query Match: 8.83% Indels: 10
DB: 2 Gaps: 2

US-09-856-221-1 (1-449) x US-08-840-683-9 (1-216)
QY 151 GCGAAAGTACTGCAAAATTCGTTAAAC-----GCTTTGTTC 186
Db 100 AlahisargThrSerLysLeuThrglnSerGlyProSerLeuProGlnAlaThrSer 119
QY 187 CTTCTCAGATAAACAACAAAGTACTGCGAGACGTTAGCAGACGCCCTATAT 246
Db 120 IleProAlaIleLeuLysAlaLeuGlnAspGluTrpAspAlaValMetLeuHisSerPhe 139
QY 247 AACTTACGTCAATCTGACAAATTCATGCGCGTGTGCATACCCATCTATCGGACA 306
Db 140 ThrLeuArgGlnGlnLeuGlnThrArgGlnGlnLeuSerHisAlaLeuTyrglnHis 159
QY 307 CCAGCAGATCCGTCCTACTG-----CTTAGTCTGCCGTACCGCCTCAAA 354
Db 160 AspAlaAlaCysArgValIleAlaArgLeuThryLysGluValThrAlaAlaArg 177

RESULT 8
US-08-555-722-9
; Sequence 9, Application US/08555722
; Patent No. 5989804
; GENERAL INFORMATION:
; APPLICANT: Androphy, Elliot J.
; APPLICANT: Chen, Jason J.
; TITLE OF INVENTION: E6-BINDING PROTEINS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII (text)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/555,722
; FILING DATE: 14-NOV-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Louis
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: NEP-003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 216 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: N-terminal
US-08-555-722-9

```

```

Alignment Scores:
Pred. No.: 1.51 Length: 216
Score: 70.00 Matches: 25
Percent Similarity: 43.59% Conservative: 9
Best Local Similarity: 32.05% Mismatches: 34
Query Match: 8.83% Indels: 10
DB: 2 Gaps: 2

US-09-856-221-1 (1-449) x US-08-555-722-9 (1-216)
QY 151 GCGAAAGTACTGCAAAATTCGTTAAAC-----GCTTTGTTC 186
Db 100 AlahisargThrSerLysLeuThrglnSerGlyProSerLeuProGlnAlaThrSer 119
QY 187 CTTCTCAGATAAACAACAAAGTACTGCGAGACGTTAGCAGACGCCCTATAT 246
Db 120 IleProAlaIleLeuLysAlaLeuGlnAspGluTrpAspAlaValMetLeuHisSerPhe 139
QY 247 AACTTACGTCAATCTGACAAATTCATGCGCGTGTGCATACCCATCTATCGGACA 306
Db 140 ThrLeuArgGlnGlnLeuGlnThrArgGlnGlnLeuSerHisAlaLeuTyrglnHis 159
QY 307 CCAGCAGATCCGTCCTACTG-----CTTAGTCTGCCGTACCGCCTCAAA 354
Db 160 AspAlaAlaCysArgValIleAlaArgLeuThryLysGluValThrAlaAlaArg 177

RESULT 9
US-09-384-301-9
; Sequence 9, Application US/09384301
; Patent No. 6296853
; GENERAL INFORMATION:
; APPLICANT: Androphy, Elliot J.
; APPLICANT: Chen, Jason J.
; TITLE OF INVENTION: E6-BINDING PROTEINS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII (text)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/384,301
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/555,722
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Louis
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: NEP-003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 216 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: N-terminal
US-09-384-301-9
Alignment Scores:
Pred. No.: 1.51 Length: 216
Score: 70.00 Matches: 25

```



Db 139 GlnIleThrValLysValSerAlaSerLeuAspGlyLys----- 151  
QY 178 GCTTTGTTCTCTCTCAGATAAACAATAAAGTTACTTGCAGACG----- 228  
Db 152 -----GlnAlaAsnAspAsnGlyGlnSerGlnTrpIleThrAsnLysGlu 166  
QY 229 TTAGCACACCCCTATATAACTTA-----CGTCATAATCTGCAATTTGATGTCAG--- 279  
Db 167 ValLysGlnAspValTyrLysLeuArgHisArgHisAspAlaValLeuThrGlyArg 186  
QY 280 -----CCGTTGTTCATTACCCATCTATCGACA 306  
Db 187 ThrValGluLeuAspAspProGlnTyrThr 197

RESULT 12  
US-09-369-700-2  
; Sequence 2, Application US/09369700  
; Patent No. 6280735  
; GENERAL INFORMATION:  
; APPLICANT: Palmer, Leslie M.  
; TITLE OF INVENTION: No. 6280735el ribg  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dechert Price & Rhoads  
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre  
; CITY: Philadelphia  
; STATE: PA  
; COUNTRY: US  
; ZIP: 19103  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/369,700  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/978,456  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Dickinson, Todd Q  
; REGISTRATION NUMBER: 28,354  
; REFERENCE/DOCKET NUMBER: P50444-9  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 215-994-2252  
; TELEFAX: 215-994-2222  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 342 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-09-369-700-2

Alignment Scores:  
Pred. No.: 2.15 Length: 342  
Score: 69.50 Matches: 28  
Percent Similarity: 42.86% Conservative: 11  
Best Local Similarity: 30.77% Mismatches: 25  
Query Match: 8.76% Indels: 25  
DB: 4 Gaps: 5

US-09-856-221-1 (1-449) x US-09-369-700-2 (1-342)

QY 73 GAAGAGCGCGCAGTCAACCACTCGTGATCATTAATGCGCAACTGCCT 132  
Db 119 AspGluArgAlaSerGlnLeuTyrGlnAspPhePheLysAlaLysGlnLeuPro 138  
QY 133 CAA-----CCGCGCGCTTCCGCGCAACCTACTGCAAAATTCGTTAAC 177  
|||

Db 139 GlnIleThrValLysValSerAlaSerLeuAspGlyLys----- 151  
QY 178 GCTTTGTTCTCTCTCAGATAAACAATAAAGTTACTTGCAGACG----- 228  
Db 152 -----GlnAlaAsnAspAsnGlyGlnSerGlnTrpIleThrAsnLysGlu 166  
QY 229 TTAGCACACCCCTATATAACTTA-----CGTCATAATCTGCAATTTGATGTCAG--- 279  
Db 167 ValLysGlnAspValTyrLysLeuArgHisArgHisAspAlaValLeuThrGlyArg 186  
QY 280 -----CCGTTGTTCATTACCCATCTATCGACA 306  
Db 187 ThrValGluLeuAspAspProGlnTyrThr 197

RESULT 13  
US-08-444-083-8  
; Sequence 8, Application US/08444083  
; Patent No. 5571675  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Joffre  
; APPLICANT: Chien, Kenneth  
; APPLICANT: King, Kathleen  
; APPLICANT: Pennica, Diane  
; APPLICANT: Wood, William  
; TITLE OF INVENTION: Cardiac Hypertrophy Factor and Uses Therefor  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genentech, Inc.  
; STREET: 460 Point San Bruno Blvd  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: patin (Genentech)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/444,083  
; FILING DATE: 17-MAY-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/233609  
; FILING DATE: 25-APR-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/286304  
; FILING DATE: 05-AUG-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Torchia, Timothy E.  
; REGISTRATION NUMBER: 36,700  
; REFERENCE/DOCKET NUMBER: 894PLD5  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415/225-8674  
; TELEFAX: 415/952-9881  
; TELEX: 910/371-7168  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 201 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
US-08-444-083-8

Alignment Scores:  
Pred. No.: 1.98 Length: 201  
Score: 39.00 Matches: 25  
Percent Similarity: 38.78% Conservative: 13  
Best Local Similarity: 25.51% Mismatches: 31  
Query Match: 8.70% Indels: 29  
DB: 1 Gaps: 3

US-09-856-221-1 (1-449) x US-08-444-083-8 (1-201)

QY 61 GAACCCGTTTGAAGAAGCCGCCAGTCAAAACCATTCGTGATCATTTATCAGCATAAATG 120  
Db 10 AspProGlnThrAspSerValSer-----LeuLeuProHisLeuGluAlaLysIle 27  
QY 121 CGGCAACTGCGTCAACGCGCGCTTGGCGGGAACGCTACTGCAATTCGTTAACCGCT 180  
Db 28 ArgGln-----ThrHisSerLeuAlaHis 35  
QY 181 TTGTTCTTCTCAGATAAACAAGTTACTGGCAGACGTTTAGCACAAACG 240  
Db 36 LeuLeuThrLysTyraLaGluGlnLeuGluTyraValGln----- 50  
QY 241 CTATATACTTACGTCAATCTGCAATTCGACGCGTGTTCATTAACCATCTAT 300  
Db 51 -----LeuGlnGlyAspProPheGlyLeuProSerPhe 61  
QY 301 GGCACACGACGATCGTCTGCTTACTGCTGCTGCGCTACCGCTCAC 352  
Db 62 SerProProArgLeuProValAlaGlyLeuSerAla-ProAlaProSerHis 78  
RESULT 14  
US-08-286-304-8  
; Sequence 8, Application US/08286304  
; Patent No. 5571893  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Joffre  
; APPLICANT: Chien, Kenneth  
; APPLICANT: King, Kathleen  
; APPLICANT: Pennica, Diane  
; APPLICANT: Wood, William  
; TITLE OF INVENTION: Cardiac Hypertrophy Factor and Uses Therefor  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genentech, Inc.  
; STREET: 460 Point San Bruno Blvd  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: patin (Genentech)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/286,304  
; FILING DATE: 05-AUG-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/233609  
; FILING DATE: 25-APR-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Hasak, Janet E.  
; REGISTRATION NUMBER: 28,616  
; REFERENCE/DOCKET NUMBER: 894P1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415/225-1896  
; TELEFAX: 415/952-9881  
; TELEX: 910/371-7168  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 201 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
US-08-286-304-8

Alignment Scores:  
Pred. No.: 1.98 Length: 201  
Score: 69.00 Matches: 25  
Percent Similarity: 38.78% Conservative: 13  
Best Local Similarity: 25.51% Mismatches: 31  
Query Match: 8.70% Indels: 29  
DB: 1 Gaps: 3

US-09-856-221-1 (1-449) x US-08-286-304-8 (1-201)  
QY 61 GAACCCGTTTGAAGAAGCCGCCAGTCAAAACCATTCGTGATCATTTATCAGCATAAATG 120  
Db 10 AspProGlnThrAspSerValSer-----LeuLeuProHisLeuGluAlaLysIle 27  
QY 121 CGGCAACTGCGTCAACGCGCGCTTGGCGGGAACGCTACTGCAATTCGTTAACCGCT 180  
Db 28 ArgGln-----ThrHisSerLeuAlaHis 35  
QY 181 TTGTTCTTCTCAGATAAACAAGTTACTGGCAGACGTTTAGCACAAACG 240  
Db 36 LeuLeuThrLysTyraLaGluGlnLeuGluTyraValGln----- 50  
QY 241 CTATATACTTACGTCAATCTGCAATTCGACGCGTGTTCATTAACCATCTAT 300  
Db 51 -----LeuGlnGlyAspProPheGlyLeuProSerPhe 61  
QY 301 GGCACACGACGATCGTCTGCTTACTGCTGCTGCGCTACCGCTCAC 352  
Db 62 SerProProArgLeuProValAlaGlyLeuSerAla-ProAlaProSerHis 78  
RESULT 15  
US-08-442-745-8  
; Sequence 8, Application US/08442745  
; Patent No. 5624806  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Joffre  
; APPLICANT: Chien, Kenneth  
; APPLICANT: King, Kathleen  
; APPLICANT: Pennica, Diane  
; APPLICANT: Wood, William  
; TITLE OF INVENTION: Cardiac Hypertrophy Factor and Uses Therefor  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genentech, Inc.  
; STREET: 460 Point San Bruno Blvd  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: patin (Genentech)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/442,745  
; FILING DATE: 17-may-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/233609  
; FILING DATE: 25-APR-1994  
; APPLICATION NUMBER: 08/286304  
; FILING DATE: 05-AUG-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Torchia, Timothy E.  
; REGISTRATION NUMBER: 36,700  
; REFERENCE/DOCKET NUMBER: 894P1D1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415/225-8674  
; TELEFAX: 415/952-9881  
; TELEX: 910/371-7168  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 201 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
US-08-442-745-8  
Alignment Scores:

```
Pred. No.:      1.98      Length:      201
Score:          69.00      Matches:      25
Percent Similarity: 38.78%      Conservative: 13
Best Local Similarity: 25.51%      Mismatches: 31
Query Match:      8.70%      Indels:      29
DB:              1          Gaps:        3

US-09-856-221-1 (1-449) x US-08-442-745-8 (1-201)

Qy  61  GAACCCGTTTAGAAGAGCCGCCAGTCAAAACCATTCGTATCATATTCAGCATAAATG 120
Db  10  AspProGlnThrAspSerValSer-----LeuLeuProHisLeuGluAlaLysile 27
Qy  121 CGGCAACTGGCTCAACGCGCGCTTGGCGGAAACGTAAGTCAAAATTCGTTAACCGCT 180
Db  28  ArgGln-----ThrHisSerLeuAlaHis 35
Qy  181 TTGTTCTCTCTCAGATAAACAAAAAACTGCAAAGTTACTGGCAGAGCTTAGCACACGC 240
Db  36  LeuLeuThrLysTyrAlaGluGlnLeuLeuGlnGluTyrValGln----- 50
Qy  241 CTATATAACTTACGTCATAATCTGACAAATTGATGGTCAGCGCTGTGTCATTACCCATCTAT 300
Db  51  -----LeuGlnGlyAspProPheGlyLeuProSerPhe 61
Qy  301 GCGACACACAGATCGCTCCGTACTGCTTAGTGTGCTCCGCTACCGCCTCAC 352
Db  62  SerProProArgLeuProValAlaGlyLeuSerAla-ProAlaProSerHis 78
```

Search completed: January 15, 2003, 15:57:42  
Job time : 21.4662 secs







```
Query Match: 54.79% Indels: 3
DB: 10 Gaps: 1
US-09-856-221-1 (1-449) x US-09-817-514A-2 (1-2516)
QY 13 CACCTATTTGGGCGATACCTTATTTTCATTGGATACGATTTGGTCAGAACCCGCTTFA 72
|||||:|||||
DB 1888 HisLeuLeuGlyAspLysProTyrLeuProLeuSerThrThrTrpSerAspProArgLeu 1907
|||||:|||||
QY 73 GAAGAAGCGCCGAGTCAAAACCATTCGTGATCATTAATCAGCAATAAATGCGCAACGCGT 132
|||||:|||||
DB 1908 AspArgAlaAlaAspIleThrThrGlnAsnAlaHisAspSerAlaIleValAlaLeuArg 1927
|||||:|||||
QY 133 CACGCGGCGGCTTGGCGGGA-----CCTACTGCMAATTCGTTAACCGCTTTG 183
|||||:|||||
DB 1928 GlnAsnIleProThrProAlaProLeuSerLeuArgSerAlaAsnThrLeuThrAspLeu 1947
|||||:|||||
QY 184 TTCCTCTCAGATAAACAACAACTGCAAGCTTACTGGCAGACGTTAGCACACGCTTA 243
|||||:|||||
DB 1948 PheLeuProGlnIleAsnGluValMetMetAsnTyrTrpGlnThrLeuAlaGlnArgVal 1967
|||||:|||||
QY 244 TATTAACCTACGTATATCTGACATTTGATGTGACCGCTTGTCTATTAACCATCTATGCG 303
|||||:|||||
DB 1968 TyrAsnLeuArgHisAsnLeuSerIleAspGlyGlnProLeuTyrLeuProIleTyrAla 1987
|||||:|||||
QY 304 ACACGAGCAGATCCGCTACTGCTTAGTGTGCTGCGCTCACCGCTCACAGGCGGAGG 363
|||||:|||||
DB 1988 ThrProAlaAspProLysAlaLeuLeuSerAlaAlaValAlaThrSerGlnGlyGlyGly 2007
|||||:|||||
QY 364 GATTTGCTCGCAGCAGTAATGCGGATGTACCGTTTCCGATTATTTCTGGAATGCCAAG 423
|||||:|||||
DB 2008 LysLeuProGluSerPheMetSerLeuTrpArgPheProHisMetLeuGluAsnAlaArg 2027
|||||:|||||
QY 424 TGGGGAGTGACCCCACTGATACAG 447
|||||:|||||
DB 2028 GlyMetValSerGlnLeuThrGln 2035
|||||:|||||
RESULT 2
US-09-817-514A-8
; Sequence 8, Application US/09817514A
; Patent No. US20020078478A1
; GENERAL INFORMATION:
; APPLICANT: ffranch-Constant, Richard
; APPLICANT: Bowen, David
; APPLICANT: Rocheleau, Thomas
; APPLICANT: Waterfield, Nicholas
; TITLE OF INVENTION: DNA SEQUENCES FROM PHOTORHABDUS LUMINESCENS
; FILE REFERENCE: 61645
; CURRENT APPLICATION NUMBER: US/09/817,514A
; PRIOR FILING DATE: 2000-03-26
; PRIOR APPLICATION NUMBER: US 60/191806
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
; LENGTH: 2504
; TYPE: PRT
; ORGANISM: Photorhabdus luminescens
US-09-817-514A-8
Alignment Scores:
Pred. No.: 6,32e-43 Length: 2504
Score: 415.00 Matches: 90
Percent Similarity: 71.33% Conservative: 17
Best Local Similarity: 60.00% Mismatches: 41
Query Match: 52.33% Indels: 3
DB: 10 Gaps: 1
US-09-856-221-1 (1-449) x US-09-817-514A-8 (1-2504)
QY 5 CAGGCACTCACCTTATTTGGCGATAA-CCTTATTTTTCATTGGATACGATTTGGTCAGAA 63
|||||:|||||
DB 1880 GlnAlaLeuAsnLeuLeuGlyAspGluProGlnValMetLeuSerThrThrTrpAlaAsn 1899
|||||:|||||
```

```

Db 1252 ArgSerSerArgLeuAlaTyrArgThrAlaAlaAlaSerGlnValLeuPheTyrProSer 1271
QY 196 ATAACAAACAAAGTACTGCGAGACGTTAGCACACACGCTATATAACTTACGT 255
Db 1272 TyrLysLys
QY 256 CATAATCTGAAATTCATGCTCAGCGCTGTCTATTACCATCTATCGCACACACGAGAT 315
Db 1275 -----ThrLysProGlyGlnProThrGlyTyrProGlnTyrAlaGluAlaLeuasp 1291
QY 316 CCGTCCCTACTGTTAGTCTGCTCCCTACCCCTCAACAGCGGAGGATTTGCTCGG 375
Db 1292 ProProLeuAlaThrGlyAsnAlaAlaAlaTyrTyrGlnGlnGlnGlnLeuArgArg 1311
RESULT 4
US-10-151-736-4
; Sequence 4, Application US/10151736
; Publication No. US20020192160A1
; GENERAL INFORMATION:
; APPLICANT: Callaghan, Michelle J.
; APPLICANT: Sutherland, Lindfield
; APPLICANT: Watts, Colin K.
; TITLE OF INVENTION: No. US20020192160A1 Human Tumour Suppressor Gene
; FILE REFERENCE: RICE-010CON
; CURRENT APPLICATION NUMBER: US/10/151,736
; CURRENT FILING DATE: 2002-08-15
; PRIOR APPLICATION NUMBER: 09/403,402
; PRIOR FILING DATE: 1999-10-21
; PRIOR APPLICATION NUMBER: PCT/AU98/00280
; PRIOR FILING DATE: 1998-04-20
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 2799
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-151-736-4
Alignment Scores:
Pred. No.: 0.898 Length: 2799
Score: 76.00 Matches: 33
Percent Similarity: 43.55% Conservative: 18
Best Local Similarity: 28.21% Mismatches: 46
Query Match: 9.58% Indels: 20
Gaps: 5
US-09-856-221-1 (1-449) x US-10-151-736-4 (1-2799)
QY 43 TTGATAACGATTGGTGCAGACCCCGTTTAGAAGAGCCCGCCAGTCAACACCATTCGTGAT 102
Db 1409 LeuGlnAsnLysTyrThrProGlyArgArgGluGluAlaIleAlaValThr----- 1425
QY 103 CATTATCAGCATAAATGCGGCACTGCTCAACGCGG-----GCCTTGCGG 150
Db 1426 -----MetArgPheLeuArgSerValAlaArgValPheValIleLeuSer 1440
QY 151 GCGAAGGTACTGCAAAATTCGTTAACCCTTTGTTCTCTCAGATAAACAACAACTG 210
Db 1441 ValGluMetAlaSerSerLysLysAsnAsnPhelProGlnProIleGlyLysCys 1460
QY 211 CAAAGTTACTGCGACAGCTTA-----GCACAACGCTATATACITACGTCTAT 258
Db 1461 LysArgValPheGlnAlaLeuLeuProTyrAlaValGluGluCysAsnValAlaGlu 1480
QY 259 AATCTGCAATATGATGGTCAGCGGTGTCTATCCCATCTATCGCACACGAGATCGG 318
Db 1481 SerLeuIleVal-----ProValArgMetGlyIle---AlaArgProThrAlaPro 1496
QY 319 TCCGTACTGCTTAGTCTGCCCTCACCCTCAACAGCGGAGGGATTTG 369
Db 1497 PheThrLeuAlaSerThrIleAspAlaMetGlnGlySerGluGluLeu 1513
RESULT 5

```

```

US-09-933-638A-12
; Sequence 12, Application US/09933638A
; Patent No. US20020160952A1
; GENERAL INFORMATION:
; APPLICANT: Kazantsev, Alexsey G.
; APPLICANT: Thompson, David E.
; APPLICANT: Housman, David E.
; TITLE OF INVENTION: INHIBITION OF PROTEIN-PROTEIN INTERACTION
; FILE REFERENCE: 01997-289001
; CURRENT APPLICATION NUMBER: US/09/933,638A
; CURRENT FILING DATE: 2001-08-20
; PRIOR APPLICATION NUMBER: US 60/226,502
; PRIOR FILING DATE: 2000-08-18
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 338
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-933-638A-12
Alignment Scores:
Pred. No.: 0.783 Length: 338
Score: 73.50 Matches: 23
Percent Similarity: 42.72% Conservative: 21
Best Local Similarity: 22.33% Mismatches: 46
Query Match: 9.27% Indels: 13
Gaps: 2
US-09-856-221-1 (1-449) x US-09-933-638A-12 (1-338)
QY 73 GAAGAAGCCCGCAGTCAACACCATTCGTGATCATTAATCGCAACTGCGT 132
Db 73 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 92
QY 133 CAACGCGCGCTTCCGCGGAAACGTACTGCAAAATTCGTTAACCGCTTTGTTCCCT 192
Db 93 GlnGlnGlnAlaValAlaAlaAlaValGlnGlnSerThrSer----- 107
QY 193 CAGATAACAAACAACTGCAAGATTTACTGGCAGACGTTAGCACACGCTATATAACTTA 252
Db 108 -----GlnGlnAlaThrGlnGlyThrSerGlyGlnAlaProGlnLeu 121
QY 253 CGTCATATCTGACAAATTCGATGTCGCGGTG-----TCATTACCATCTATCGGACA 306
Db 122 PheHisSerGlnThrLeuThrThrAlaProLeuProGlyThrThrProLeuTyrProSer 141
QY 307 CCACGACATCCGTCCTACTGCTTAGTCTGCTCACCCTCACACGCGGAGGGAT 366
Db 142 ProMetThrProMetThrProIleThrProAlaThrProAlaSerGluSerGlyIle 161
QY 367 TTGCCTCGG 375
Db 162 ValProGln 164
RESULT 6
US-09-849-243-16
; Sequence 16, Application US/09849243
; Patent No. US20020157127A1
; GENERAL INFORMATION:
; APPLICANT: Kirschbaum, Bernd
; APPLICANT: Berglund, Erick
; APPLICANT: Meisterernst, Michael
; APPLICANT: Polites, Greg
; TITLE OF INVENTION: PURIFICATION OF HIGHER ORDER TRANSGENIC
; COMPLEXES FROM TRANSGENIC
; NON-HUMAN ANIMALS
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HELLER, EHRMAN, WHITE & MCAULIFFE
; STREET: 1666 K Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.

```

COUNTRY: USA  
 ZIP: 20006  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/849,243  
 FILING DATE: 07-May-2001  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Granados, Patricia D.  
 REGISTRATION NUMBER: 33,683  
 REFERENCE/DOCKET NUMBER: 38005-0148  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (202)912-2000  
 TELEFAX: (202)912-2020

INFORMATION FOR SEQ ID NO: 16:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 371 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear

MOLECULE TYPE: protein  
 SEQUENCE DESCRIPTION: SEQ ID NO: 16:  
 US-09-849-243-16

Alignment Scores:  
 Pred. No.: 0.813 Length: 371  
 Score: 73.50 Matches: 23  
 Percent Similarity: 42.72% Conservative: 21  
 Best Local Similarity: 22.33% Mismatches: 46  
 Query Match: 9.27% Indels: 13  
 DB: 9 Gaps: 2

US-09-856-221-1 (1-449) x US-09-849-243-16 (1-371)

QY 73 GAAGAGCGCGCAGTCAACCATTCGTGATCATTAACAGATAAAATCGGCAACTGCCT 132  
 Db 105 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 124  
 QY 133 CAACGCGCGCGTTCGCGCGGCAACGCTACTGCAAAATCGTTACCGCTTTGTCCTTCCT 192  
 Db 125 GlnGlnGlnAlaValAlaAlaAlaValGlnGlnSerThrSer----- 139  
 QY 193 CAGATAAAACAAACTGCAAGGTACTGCGCAGCGTTAGCACACGCCCTATATACTTA 252  
 Db 140 -----GlnGlnAlaThrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnLeu 153  
 QY 253 CGTCATATCTGCAATGATGTCAGCGGTG-----TCATTACCATCTATGCGACA 306  
 Db 154 PheHisSerGlnThrLeuThrThrAlaProLeuProGlnGlnThrProLeuThrProSer 173  
 QY 307 CCAGCAGATCGCGTCTGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 366  
 Db 174 ProMetThrProMetThrProLeuThrProAlaThrProAlaSerGluSerSerGlyIle 193  
 QY 367 TCGCTCGG 375  
 Db 194 ValProGln 196

RESULT 7

US-09-944-413-55  
 ; Sequence 55, Application US/09944413  
 ; Patent No US20020156004A1

GENERAL INFORMATION:

APPLICANT: Baker, Kevin  
 APPLICANT: Botstein, David  
 APPLICANT: Eaton, Dan  
 APPLICANT: Ferrara, Napoleone  
 APPLICANT: Filvaroff, Ellen  
 APPLICANT: Gerritsen, Mary  
 APPLICANT: Goddard, Audrey  
 APPLICANT: Godowski, Paul

APPLICANT: Grimaldi, Christopher  
 APPLICANT: Gurney, Austin  
 APPLICANT: Hillan, Kenneth  
 APPLICANT: Kijavini, Ivar  
 APPLICANT: Napier, Mary  
 APPLICANT: Roy, Margaret  
 APPLICANT: Tumas, Daniel  
 APPLICANT: Wood, William  
 TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
 FILE REFERENCE: P2548P1C1  
 CURRENT APPLICATION NUMBER: US/09/944,413  
 CURRENT FILING DATE: 2001-09-26  
 PRIOR APPLICATION NUMBER: 09/866,028  
 PRIOR FILING DATE: 2001-05-25  
 PRIOR APPLICATION NUMBER: 60/067,411  
 PRIOR FILING DATE: December 3, 1997  
 PRIOR APPLICATION NUMBER: 60/069,334  
 PRIOR FILING DATE: December 11, 1997  
 PRIOR APPLICATION NUMBER: 60/069,335  
 PRIOR FILING DATE: December 11, 1997  
 PRIOR APPLICATION NUMBER: 60/069,278  
 PRIOR FILING DATE: December 11, 1997  
 PRIOR APPLICATION NUMBER: 60/069,425  
 PRIOR FILING DATE: December 12, 1997  
 PRIOR APPLICATION NUMBER: 60/069,696  
 PRIOR FILING DATE: December 16, 1997  
 PRIOR APPLICATION NUMBER: 60/069,694  
 PRIOR FILING DATE: December 16, 1997  
 PRIOR APPLICATION NUMBER: 60/069,702  
 PRIOR FILING DATE: December 16, 1997  
 PRIOR APPLICATION NUMBER: 60/069,870  
 PRIOR FILING DATE: December 17, 1997  
 PRIOR APPLICATION NUMBER: 60/069,873  
 PRIOR FILING DATE: December 17, 1997  
 PRIOR APPLICATION NUMBER: 60/068,017  
 PRIOR FILING DATE: December 18, 1997  
 PRIOR APPLICATION NUMBER: 60/070,440  
 PRIOR FILING DATE: January 5, 1998  
 PRIOR APPLICATION NUMBER: 60/074,086  
 PRIOR FILING DATE: February 9, 1998  
 PRIOR APPLICATION NUMBER: 60/074,092  
 PRIOR FILING DATE: February 9, 1998  
 PRIOR APPLICATION NUMBER: 60/075,945  
 PRIOR FILING DATE: February 25, 1998  
 PRIOR APPLICATION NUMBER: 60/112,850  
 PRIOR FILING DATE: December 16, 1998  
 PRIOR APPLICATION NUMBER: 60/113,296  
 PRIOR FILING DATE: December 22, 1998  
 PRIOR APPLICATION NUMBER: 60/146,222  
 PRIOR FILING DATE: July 28, 1999  
 PRIOR APPLICATION NUMBER: PCT/US98/19330  
 PRIOR FILING DATE: September 16, 1998  
 PRIOR APPLICATION NUMBER: PCT/US98/25108  
 PRIOR FILING DATE: December 1, 1998  
 PRIOR APPLICATION NUMBER: 09/216,021  
 PRIOR FILING DATE: December 16, 1998  
 PRIOR APPLICATION NUMBER: 09/218,517  
 PRIOR FILING DATE: December 22, 1998  
 PRIOR APPLICATION NUMBER: 09/254,311  
 PRIOR FILING DATE: March 3, 1999  
 PRIOR APPLICATION NUMBER: PCT/US99/12252  
 PRIOR FILING DATE: June 22, 1999  
 PRIOR APPLICATION NUMBER: PCT/US99/21090  
 PRIOR FILING DATE: September 15, 1999  
 PRIOR APPLICATION NUMBER: PCT/US99/28409  
 PRIOR FILING DATE: No. US20020156004A1ember 30, 1999  
 PRIOR APPLICATION NUMBER: PCT/US99/28313  
 PRIOR FILING DATE: No. US20020156004A1ember 30, 1999  
 PRIOR APPLICATION NUMBER: PCT/US99/28301  
 PRIOR FILING DATE: December 1, 1999  
 PRIOR APPLICATION NUMBER: PCT/US99/30095  
 PRIOR FILING DATE: December 16, 1999

```

; PRIOR APPLICATION NUMBER: PCT/US00/03565
; PRIOR FILING DATE: February 11, 2000
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: February 22, 2000
; PRIOR APPLICATION NUMBER: PCT/US00/05841
; PRIOR FILING DATE: March 2, 2000
; PRIOR APPLICATION NUMBER: PCT/US00/08439
; PRIOR FILING DATE: March 30, 2000
; PRIOR APPLICATION NUMBER: PCT/US00/14042
; PRIOR FILING DATE: May 22, 2000
; PRIOR APPLICATION NUMBER: PCT/US00/20710
; PRIOR FILING DATE: July 28, 2000
; PRIOR APPLICATION NUMBER: PCT/US00/32678
; PRIOR FILING DATE: December 1, 2000
; PRIOR APPLICATION NUMBER: PCT/US01/06520
; PRIOR FILING DATE: February 28, 2001
; NUMBER OF SEQ ID NOS: 120
; SEQ ID NO 55
; LENGTH: 694
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-944-413-55

Alignment Scores:
Pred. No.: 1.86
Score: 71.50
Percent Similarity: 45.54%
Best Local Similarity: 24.11%
Query Match: 9.02%
DB: 9

US-09-856-221-1 (1-449) x US-09-944-413-55 (1-694)
QY 58 TCAGACCCGTTTAGAGAGCCGCCGATCAACCATTCGTGATCATATTACGATATA 117
Db 118 SerGlyAspArgVallysluylsArgAsnlysthrThrGluGlnsluGlyGly 137
QY 118 ATGCGGCACTGCTCAACCGCGCGCTTGGCGGCAAACTGTCGAAATCGTTAAC 177
Db 138 ThrGluilePheArgAlaSerAlaValIleProSerlyslsAla----- 153
QY 178 GCTTGTGCTCTCTCTCAGATAAACAACAACTGCAAAAGTACTGGCAGAGTACGACAA 237
Db 154 AlaPheLeu-----SeryrGluGluLeuLeuGlnArg 165
QY 238 GCGCTATATACCTAGTCATATCTGACATTTGATGTCAGCGGCTGTGTCATCCCATC 297
Db 166 ArgLeuGlyLysrGluHisSerIleSerValArgProGlnGlnLeuSerGlyArgLeu 185
QY 298 TATCGCACACGACGATCCGTCCTGCTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 357
Db 186 -----SerValAspValAsnleLeuGluSerAlaGlyleAlaSerLeuGlu--- 201
QY 358 GGAGGGATTTGCTCGGACAGTATGCGGATGTAC 393
Db 202 -----ValleuProLeuHis 206

RESULT 8
US-09-944-403-55
; Sequence 55, Application US/09944403
; Patent No. US20020165143A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin
; APPLICANT: Botstein, David
; APPLICANT: Eaton, Dan
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gottisen, Mary
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul
; APPLICANT: Grimaldi, Christopher
; APPLICANT: Gurney, Austin
; APPLICANT: Hillan, Kenneth

```

```

; APPLICANT: Kijavini, Ivar
; APPLICANT: Napier, Mary
; APPLICANT: Roy, Margaret
; APPLICANT: Tamas, Daniel
; APPLICANT: Wood, William
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P2548P1C1
; CURRENT APPLICATION NUMBER: US/09/944,403
; CURRENT FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 09/866,028
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: 60/067,411
; PRIOR FILING DATE: December 3, 1997
; PRIOR APPLICATION NUMBER: 60/069,334
; PRIOR FILING DATE: December 11, 1997
; PRIOR APPLICATION NUMBER: 60/069,335
; PRIOR FILING DATE: December 11, 1997
; PRIOR APPLICATION NUMBER: 60/069,278
; PRIOR FILING DATE: December 11, 1997
; PRIOR APPLICATION NUMBER: 60/069,425
; PRIOR FILING DATE: December 12, 1997
; PRIOR APPLICATION NUMBER: 60/069,696
; PRIOR FILING DATE: December 16, 1997
; PRIOR APPLICATION NUMBER: 60/069,694
; PRIOR FILING DATE: December 16, 1997
; PRIOR APPLICATION NUMBER: 60/069,702
; PRIOR FILING DATE: December 16, 1997
; PRIOR APPLICATION NUMBER: 60/069,870
; PRIOR FILING DATE: December 17, 1997
; PRIOR APPLICATION NUMBER: 60/069,873
; PRIOR FILING DATE: December 17, 1997
; PRIOR APPLICATION NUMBER: 60/068,017
; PRIOR FILING DATE: December 18, 1997
; PRIOR APPLICATION NUMBER: 60/070,440
; PRIOR FILING DATE: January 5, 1998
; PRIOR APPLICATION NUMBER: 60/074,086
; PRIOR FILING DATE: February 9, 1998
; PRIOR APPLICATION NUMBER: 60/074,092
; PRIOR FILING DATE: February 9, 1998
; PRIOR APPLICATION NUMBER: 60/075,945
; PRIOR FILING DATE: February 25, 1998
; PRIOR APPLICATION NUMBER: 60/112,850
; PRIOR FILING DATE: December 16, 1998
; PRIOR APPLICATION NUMBER: 60/113,296
; PRIOR FILING DATE: December 22, 1998
; PRIOR APPLICATION NUMBER: 60/146,222
; PRIOR FILING DATE: July 28, 1999
; PRIOR APPLICATION NUMBER: PCT/US98/19330
; PRIOR FILING DATE: September 16, 1998
; PRIOR APPLICATION NUMBER: PCT/US98/25108
; PRIOR FILING DATE: December 1, 1998
; PRIOR APPLICATION NUMBER: 09/216,021
; PRIOR FILING DATE: December 16, 1998
; PRIOR APPLICATION NUMBER: 09/218,517
; PRIOR FILING DATE: December 22, 1998
; PRIOR APPLICATION NUMBER: 09/254,311
; PRIOR FILING DATE: March 3, 1999
; PRIOR APPLICATION NUMBER: PCT/US99/12252
; PRIOR FILING DATE: June 22, 1999
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: September 15, 1999
; PRIOR APPLICATION NUMBER: PCT/US99/28409
; PRIOR FILING DATE: No. US20020165143A1eember 30, 1999
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: No. US20020165143A1eember 30, 1999
; PRIOR APPLICATION NUMBER: PCT/US99/28301
; PRIOR FILING DATE: December 1, 1999
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: December 16, 1999
; PRIOR APPLICATION NUMBER: PCT/US00/03565
; PRIOR FILING DATE: February 11, 2000
; PRIOR APPLICATION NUMBER: PCT/US00/04414

```

[illegible]

```

; PRIOR APPLICATION NUMBER: PCT/US00/14042
; PRIOR FILING DATE: May 22, 2000
; PRIOR APPLICATION NUMBER: PCT/US00/20710
; PRIOR FILING DATE: July 28, 2000
; PRIOR APPLICATION NUMBER: PCT/US00/32678
; PRIOR FILING DATE: December 1, 2000
; PRIOR APPLICATION NUMBER: PCT/US01/06520
; PRIOR FILING DATE: February 28, 2001
; NUMBER OF SEQ ID NOS: 120
; SEQ ID NO 55
; LENGTH: 694
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-944-896-55

Alignment Scores:
Pred. No.: 1.86 Length: 694
Score: 71.50 Matches: 27
Percent Similarity: 45.54% Conservative: 24
Best Local Similarity: 24.11% Mismatches: 38
Query Match: 9.02% Indels: 23
DB: 9 Gaps: 4

US-09-856-221-1 (1-449) x US-09-944-896-55 (1-694)
QY 58 TCAGACCCCGTTTACGAGAGCCGAGTCAACACATTGCTGATCATTTATCAGCATAAA 117
Db 118 SerGlyAspArgVallyGluGlyAsnLysThrThrGluGluAsnGlyGlyGly 137
QY 118 ATGGCGCACTGCTCAAGCGCGCCCTTGCAGGAAAGCTACTGCAAAATTCGTTAAC 177
Db 138 ThrGluilePheArgAlaSerAlaValileProSerLysAspLysAla----- 153
QY 178 GCITTTGCTCTCTCTCAGATCAACAAAAGTCTGCGCAGAGCTTACGACAA 237
Db 154 AlaPheLeu-----SerTyrGluLeuLeuGlnArg 165
QY 238 CGCTATATAACTTACGTATATCTGACAAATGATGTCGAGCGGTGTCTATCCCATC 297
Db 166 ArgLeuGlyLysTyrGluHisSerileSerValArgProGlnGlnLeuSerGlyArgLeu 185
QY 298 TATCGACACACGACGATCGCTCTGCTAGTCTGCTGCTGCTGCTGCTGCTGCTGCT 357
Db 186 -----SerValAspValAsnLeuGluSerAlaGlyIleAlaSerLeuGlu--- 201
QY 358 GGAGGGGATGTCCTCGGACAGATGATCCGATGATC 393
Db 202 -----ValleuProLeuHis 206

RESULT 10
US-09-944-944-55
; Sequence 55, Application US/09944944
; Patent No. US20020173463A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin
; APPLICANT: Botstein, David
; APPLICANT: Eaton, Dan
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gerritsen, Mary
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul
; APPLICANT: Grimaldi, Christopher
; APPLICANT: Gurney, Austin
; APPLICANT: Hillan, Kenneth
; APPLICANT: Kljavin, Ivar
; APPLICANT: Napier, Mary
; APPLICANT: Roy, Margaret
; APPLICANT: Tomas, Daniel
; APPLICANT: Wood, William
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P2548P1C1
; CURRENT APPLICATION NUMBER: US/09/944,944
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 09/866,028
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: 60/067,411
; PRIOR FILING DATE: December 3, 1997
; PRIOR APPLICATION NUMBER: 60/069,334
; PRIOR FILING DATE: December 11, 1997
; PRIOR APPLICATION NUMBER: 60/069,335
; PRIOR FILING DATE: December 11, 1997
; PRIOR APPLICATION NUMBER: 60/069,278
; PRIOR FILING DATE: December 11, 1997
; PRIOR APPLICATION NUMBER: 60/069,425
; PRIOR FILING DATE: December 12, 1997
; PRIOR APPLICATION NUMBER: 60/069,696
; PRIOR FILING DATE: December 16, 1997
; PRIOR APPLICATION NUMBER: 60/069,694
; PRIOR FILING DATE: December 16, 1997
; PRIOR APPLICATION NUMBER: 60/069,702
; PRIOR FILING DATE: December 16, 1997
; PRIOR APPLICATION NUMBER: 60/069,870
; PRIOR FILING DATE: December 17, 1997
; PRIOR APPLICATION NUMBER: 60/069,873
; PRIOR FILING DATE: December 17, 1997
; PRIOR APPLICATION NUMBER: 60/068,017
; PRIOR FILING DATE: December 18, 1997
; PRIOR APPLICATION NUMBER: 60/070,440
; PRIOR FILING DATE: January 5, 1998
; PRIOR APPLICATION NUMBER: 60/074,086
; PRIOR FILING DATE: February 9, 1998
; PRIOR APPLICATION NUMBER: 60/074,092
; PRIOR FILING DATE: February 9, 1998
; PRIOR APPLICATION NUMBER: 60/075,945
; PRIOR FILING DATE: February 25, 1998
; PRIOR APPLICATION NUMBER: 60/112,850
; PRIOR FILING DATE: December 16, 1998
; PRIOR APPLICATION NUMBER: 60/113,296
; PRIOR FILING DATE: December 22, 1998
; PRIOR APPLICATION NUMBER: 60/146,222
; PRIOR FILING DATE: July 28, 1999
; PRIOR APPLICATION NUMBER: PCT/US98/19330
; PRIOR FILING DATE: September 16, 1998
; PRIOR APPLICATION NUMBER: PCT/US98/25108
; PRIOR FILING DATE: December 1, 1998
; PRIOR APPLICATION NUMBER: 09/216,021
; PRIOR FILING DATE: December 16, 1998
; PRIOR APPLICATION NUMBER: 09/218,517
; PRIOR FILING DATE: December 22, 1998
; PRIOR APPLICATION NUMBER: 09/254,311
; PRIOR FILING DATE: March 3, 1999
; PRIOR APPLICATION NUMBER: PCT/US99/12252
; PRIOR FILING DATE: June 22, 1999
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: September 15, 1999
; PRIOR APPLICATION NUMBER: PCT/US99/28409
; PRIOR FILING DATE: No. US20020173463A1ember 30, 1999
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: No. US20020173463A1ember 30, 1999
; PRIOR APPLICATION NUMBER: PCT/US99/28301
; PRIOR FILING DATE: December 1, 1999
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: December 16, 1999
; PRIOR APPLICATION NUMBER: PCT/US00/03565
; PRIOR FILING DATE: February 11, 2000
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: February 22, 2000
; PRIOR APPLICATION NUMBER: PCT/US00/05841
; PRIOR FILING DATE: March 2, 2000
; PRIOR APPLICATION NUMBER: PCT/US00/08439
; PRIOR FILING DATE: March 30, 2000
; PRIOR APPLICATION NUMBER: PCT/US00/14042
; PRIOR FILING DATE: May 22, 2000
; PRIOR APPLICATION NUMBER: PCT/US00/20710

```







Db 138 ThrGluIlePheArgAlaSerAlaValIleProSerLysAspLysAla----- 153  
QY 178 GCTTTGTTCCCTTCCTCAGATAAACAACAACTGCAAGTTACTGGCAGACGTTAGCACAA 237  
Db 154 AlaPhePheLeu-----SerTyTGluGluLeuLeuGlnArg 165  
QY 238 CGCCTATATAACTTACGTCAATACTGACAAATTGATGGTCAGCCGTTGTCAATTACCCATC 297  
Db 166 ArgLeuGlyLysTyTGluHisSerIleSerValArgProGlnGlnLeuSerGlyArgLeu 185  
QY 298 TATGCGACACACAGACATCCGTCGTACTGCTTAGTCGTCGCGTCACCGCCTCACAAGGC 357  
Db 186 -----SerValAspValAsnIleLeuGluSerAlaGlyIleAlaSerLeuGlu--- 201  
QY 358 GGAGGGGATTTCCTCGCACAGATTAATGCCGATGTAC 393  
Db 202 -----ValLeuProLeuHis 206

RESULT 15

US-09-866-028-55  
; Sequence 55, Application US/09866028  
; Patent No. US20020058309A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin  
; APPLICANT: Botstein, David  
; APPLICANT: Eaton, Dan  
; APPLICANT: Ferrara, Napoleone  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Gerritsen, Mary  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul  
; APPLICANT: Grimaldi, Christopher  
; APPLICANT: Gurney, Austin  
; APPLICANT: Hillan, Kenneth  
; APPLICANT: Kljavin, Ivar  
; APPLICANT: Napier, Mary  
; APPLICANT: Roy, Margaret  
; APPLICANT: Tumas, Daniel  
; APPLICANT: Wood, William  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: P2548P1C1  
; CURRENT APPLICATION NUMBER: US/09/866,028  
; CURRENT FILING DATE: 2001-05-25  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 120  
; SEQ ID NO 55  
; LENGTH: 694  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
US-09-866-028-55

Alignment Scores:  
Pred. NO.: 1.86 Length: 694  
Score: 71.50 Matches: 27  
Percent Similarity: 45.54% Conservative: 24  
Best Local Similarity: 24.11% Mismatches: 38  
Query Match: 9.02% Indels: 23  
DB: 10 Gaps: 4

US-09-856-221-1 (1-449) x US-09-866-028-55 (1-694)

QY 58 TCAGAACCCGCTTTAGAGAGCCGCCAGTCAACCATTCGTGATCATATTATACGATATAA 117  
Db 118 SerGlyAspArgValLysGluLysArgAsnLysThrGluGluAsnGlyGluLysGly 137  
QY 118 ATGGCGCACTGGTCAACCGCGCTTCCGCGGCAACCGTACTGCAAAATTCGTTAAC 177  
Db 138 ThrGluIlePheArgAlaSerAlaValIleProSerLysAspLysAla----- 153  
QY 178 GCTTTGTTCCCTTCCTCAGATAAACAACAACTGCAAGTTACTGGCAGACGTTAGCACAA 237  
Db 154 AlaPhePheLeu-----SerTyTGluGluLeuLeuGlnArg 165

QY 238 CGCCTATATAACTTACGTCAATACTGACAAATTGATGGTCAGCCGTTGTCAATTACCCATC 297  
Db 166 ArgLeuGlyLysTyTGluHisSerIleSerValArgProGlnGlnLeuSerGlyArgLeu 185  
QY 298 TATGCGACACACAGACATCCGTCGTACTGCTTAGTCGTCGCGTCACCGCCTCACAAGGC 357  
Db 186 -----SerValAspValAsnIleLeuGluSerAlaGlyIleAlaSerLeuGlu--- 201  
QY 358 GGAGGGGATTTCCTCGCACAGATTAATGCCGATGTAC 393  
Db 202 -----ValLeuProLeuHis 206

Search completed: January 15, 2003, 15:58:34  
Job time : 13.4233 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model  
Run on: January 15, 2003, 15:44:53 ; Search time 20.6921 Seconds  
(without alignments)  
4172.064 Million cell updates/sec

Title: US-09-856-221-1  
Perfect score: 793  
Sequence: 1 tgtcaggcactcacttat.....gtgaccacatgatacagtt 449

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 283224 seqs, 96134422 residues  
Total number of hits satisfying chosen parameters: 566448

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters: -MODEL=framet\_n2p\_model -DEV=xlp  
-Q=/cgn2.1/USFTO\_spool/US09856221/runat\_15012003\_153925\_12500/app\_query.fasta\_1.1948  
-DB=PIR\_73 -QFMT=fastan -SUFFIX=rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THRESHOLD=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09856221 @CGL\_1.1.95 @runat\_15012003\_153925\_12500 -NCPU=3  
-NO\_XLPY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=10 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7  
-Ygapop=10 -Ygapext=0.5 -Delop=6 -Delext=7

Database : PIR\_73: \*  
1: pir1: \*  
2: pir2: \*  
3: pir3: \*  
4: pir4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES										
Result No.	Score	Query Match	Length	DB ID	Description					
1	84.5	10.7	270	2	AB1858	tryptophan synthas				
2	84.5	10.7	568	2	C72129	probable outer mem				
3	84.5	10.7	568	2	G86493	probable leader pe				
c 4	84	10.8	317	2	AD3067	hypothetical prote				
c 5	84	10.8	338	2	F98219	hypothetical prote				
6	78	9.8	640	2	T41977	hypothetical prote				
c 7	77.5	9.9	307	2	B69837	hypothetical prote				
8	76	9.6	255	2	C69681	peptide synthetase				
9	75.5	9.5	622	2	T10009	probable serine/th				
10	75	9.5	316	1	S34437	transcription init				
c 11	75	9.6	635	2	F96660	protein F2K1.10 l				
12	74.5	9.4	407	2	A95155	hypothetical prote				
13	74.5	9.4	1437	2	T31093	probable protein-t				
14	74	9.3	1186	2	T19050	hypothetical prote				

15	73.5	9.3	297	1	I51648	transcription init
16	73.5	9.3	339	1	TWH02D	transcription init
17	73.5	9.3	450	2	AG0222	flagellar hook-len
18	73.5	9.3	622	2	H86910	probable serine/th
19	73.5	9.3	1718	1	JQ1734	genome polypeptide
20	73	9.2	277	2	S56591	hypothetical 30.5K
21	72.5	9.1	936	2	D97630	hypothetical prote
22	72.5	9.1	1078	2	PC4198	peptide synthetase
23	72.5	9.1	2688	2	I49477	alpha-A crystallin
24	72	9.1	442	2	A41015	dihydrofolamide S
25	72	9.1	488	2	T38481	probable pre-mRNA
26	72	9.1	881	2	E82097	protein-p-ii uridy
27	72	9.1	1145	2	S13643	PRP22 protein - ye
28	71.5	9.0	201	2	A30833	ribosomal protein
29	71.5	9.0	268	2	C83104	hypothetical prote
30	71.5	9.0	470	1	A25685	hypothetical prote
31	71.5	9.0	470	1	A25685	negative sexual co
32	71.5	9.0	820	2	T39796	ubiquitin-protein
33	71	9.0	275	2	AC2293	hypothetical prote
34	71	9.0	302	1	JC4059	transcription init
35	71	9.0	317	2	T39736	hypothetical prote
36	71	9.0	380	2	S75348	hypothetical prote
37	71	9.0	509	2	T16846	hypothetical prote
38	71	9.0	886	2	A85905	hypothetical prote
39	71	9.0	886	2	C65036	hypothetical prote
40	71	9.0	2441	2	S39161	CREB-binding prote
41	70.5	8.9	365	2	T24955	hypothetical prote
42	70.5	8.9	694	2	A69768	transcription anti
43	70.5	8.9	901	2	F83781	transposase (08) /
44	70	8.8	303	2	A11070	probable membrane
45	70	8.8	321	2	T47280	transposase-like p

ALIGNMENTS

RESULT 1  
AB1858  
tryptophan synthase alpha chain [imported] - Nostoc sp. (strain PCC 7120)  
C:Species: Nostoc sp.  
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 30-Jun-2002  
C:Accession: AB1858  
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Irigu  
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata  
DNA Res. 8, 205-213, 2001  
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium  
A:Reference number: AB1807; MUID:21595285; PMID:11759840  
A:Accession: AB1858  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-270 <KUR>  
A:Cross-references: GB:BA000019; PIDN:BA072369.1; PID:g17129756; GSPDB:GN00179  
A:Experimental source: strain PCC 7120  
C:Genetics:  
A:Gene: trpA  
C:Superfamily: tryptophan synthase alpha chain; tryptophan synthase alpha chain homol

ALIGNMENT SCORES										
Pred No.	Score	Length	Matches	Conservative	Mismatches	Indels	Gaps			
1	84.50	270	40	14	34	27	8			
US-09-856-221-1 (1-449) x AB1858 (1-270)										
Qy 91	ACCATTCGTCATCATATCAGCATTAATGCGCACTGCGTCAACGCCGCTTCGCG	150								
Db 3	ThrlleSerAspHisphelin---SerLeuArgGlnArgGlnGlnCysAlaLeuIlePro	21								
Qy 151	GCGAAACGTAAGTCAAAATTCGTTAACCGCTTTGTTCCTCCTCAGATAAACAACAACTG	210								

```
Db 22 -----pHeilleThraAlaGlyAspProAspLeu 30
QY 211 CAAGTACTAGCGAGCGTGTAGCAACCGCTATATACTTA---CGTCATAATCTGACA 267
Db 31 -----GlnThrThraAlaGluAlaLeuArgIleLeuAspArgHisGly----- 44
QY 268 ATTGATGCTGACCGCTGTTCATTACCATCTATCGCACCA---GCAGATCCGTCGTA 324
Db 45 AlaAspPheilleGluLeuGlyValPro---TyrSerAspProLeuAlaAspGlyProVal 63
QY 325 CTGCTTACTGCTGCTGCTCAGCGCTCACAAGCGCGAGGGGATTTGCCGCG----- 375
Db 64 IleGlnAlaAlaAlaThraArgAlaLeuGlnArgGlyThrLysLeuGluGlnValLeuGlu 83
QY 376 -----ACAGTAATGCGCGATGTACCGTTTTCGATATTCTG 411
Db 84 MetLeuGluThrValIleProSerLeuGlnAlaProIleLeu 98
RESULT 2
C72129
Probable outer membrane leader peptide (omp) CPN0021 - Chlamydothilla pneumoniae (strains
C:Species: Chlamydothilla pneumoniae, Chlamydia pneumoniae
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000
C:Accession: C72129; B81542
R:Kallman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;
Nature Genet. 21, 385-389, 1999
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A:Reference number: A72000; MUID:99206606; PMID:10192388
A:Accession: C72129
A:Molecule type: DNA
A:Residues: 1-568 <ARN>
A:Cross-References: GB:AE001587; GB:AE001363; NID:94376271; PIDN:AAD18174.1; PID:9437627
A:Experimental source: strain CWI029
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
, C.; Dodson, R.; Gwin, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,
Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.
A:Reference number: B81500; MUID:20150255; PMID:10684955
A:Accession: B81542
A:Molecule type: DNA
A:Residues: 1-568 <REA>
A:Cross-References: GB:AE002234; GB:AE002161; NID:97189667; PIDN:AAF38558.1; PID:97189667
A:Experimental source: strain AR39, HL cells
C:Genetics:
A:Gene: CPN0021; CP0755
Alignment Scores:
Pred. No.: 1.32 Length: 568
Score: 84.50 Matches: 40
Percent Similarity: 42.66% Conservative: 21
Best Local Similarity: 27.97% Mismatches: 55
Query Match: 10.66% Indels: 27
DB: 2 Gaps: 6
US-09-856-221-1 (1-449) x C72129 (1-568)
QY 58 TCAGAACCCCGTTTAGAAGAGCC-----GCCAGTCAACACCATTCGTGAT----- 102
Db 232 AlaSerProGlnAspGlnGluAlaIleLeuTyraAlaLeuGlyLysLeuLysAspGlyGln 251
QY 103 CATTATCAGCATAAATGCGCAACTGCGTCAACCGCGCGCTTGGCGGCAACGATCT 162
Db 252 SerTyrrAsnIleLysLysGlnLeuGlnLysProAspValThrLeuAlaAla 271
QY 163 GCAATTCGTTAACCCTGTTG-----TTCCCTTCCTCAGATAACAAAAA 207
Db 272 AlaGlnAlaLeuIleAlaLeuGlyLysGluAlaAspAlaLeuProValIleLysLys--- 290
QY 208 CTGCAAGTCTAGCGAGCGTTAGCACACGC-----CTATATACTTACGTCAT 258
Db 291 -----GlnAlaLeuGluAlaArgProArgAlaLeuTyraAlaLeuArgHis 305
QY 259 AATCTGACAAATTGATGTCAGCGCTTGCATTACCATCTATCGCATCGACGACGATCCG 318
Db 306 LeuProSerGluIleGlyIleProIleAlaLeuProIleLeuLysThrLysAsnSer 325
QY 319 TCCGACTGCTTAGTGTGCGGTCACCGCTCACAAGCGCGGAGGATTTGCCCT----- 372
Db 326 GluAlaLysLeuAsnValAlaLeuAlaLeuGluLeuGlyCysAspThrProLysLeu 345
QY 373 -----CGACAGTAATGCGATGTCACCGTTTCCGATATTCTCGGAA 414
Db 346 LeuGluTyrrIleThrGluArgLeuValGlnProHisTyrAsnGluThrLeuAlaLeuSer 365
QY 415 AATGCCAAG 423
Db 366 PheSerLys 368
RESULT 3
G86493
Probable leader peptide Omp [imported] - Chlamydothilla pneumoniae (strain J138)
C:Species: Chlamydothilla pneumoniae, Chlamydia pneumoniae
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 02-Mar-2001
C:Accession: G86493
R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.;
Nucleic Acids Res. 28, 2311-2314, 2000
A:Title: Comparison of whole genome sequences of Chlamydia pneumoniae J138.
A:Reference number: A86491; MUID:20330349; PMID:10871362
A:Accession: G86493
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-568 <SFO>
A:Cross-References: GB:BA000008; NID:98978396; PIDN:BAA98233.1; GSPDB:GN00142
A:Experimental source: strain J138
C:Genetics:
A:Gene: CPJ0021
Alignment Scores:
Pred. No.: 1.32 Length: 568
Score: 84.50 Matches: 40
Percent Similarity: 42.66% Conservative: 21
Best Local Similarity: 27.97% Mismatches: 55
Query Match: 10.66% Indels: 27
DB: 2 Gaps: 6
US-09-856-221-1 (1-449) x G86493 (1-568)
QY 58 TCAGAACCCCGTTTAGAAGAGCC-----GCCAGTCAACACCATTCGTGAT----- 102
Db 232 AlaSerProGlnAspGlnGluAlaIleLeuTyraAlaLeuGlyLysLeuLysAspGlyGln 251
QY 103 CATTATCAGCATAAATGCGCAACTGCGTCAACCGCGCGCTTGGCGGCAACGATCT 162
Db 252 SerTyrrAsnIleLysLysGlnLeuGlnLysProAspValThrLeuAlaAla 271
QY 163 GCAATTCGTTAACCCTGTTG-----TTCCCTTCCTCAGATAACAAAAA 207
Db 272 AlaGlnAlaLeuIleAlaLeuGlyLysGluAlaAspAlaLeuProValIleLysLys--- 290
QY 208 CTGCAAGTCTAGCGAGCGTTAGCACACGC-----CTATATACTTACGTCAT 258
Db 291 -----GlnAlaLeuGluAlaArgProArgAlaLeuTyraAlaLeuArgHis 305
QY 259 AATCTGACAAATTGATGTCAGCGCTTGCATTACCATCTATCGCATCGACGACGATCCG 318
Db 306 LeuProSerGluIleGlyIleProIleAlaLeuProIleLeuLysThrLysAsnSer 325
QY 319 TCCGACTGCTTAGTGTGCGGTCACCGCTCACAAGCGCGGAGGATTTGCCCT----- 372
Db 326 GluAlaLysLeuAsnValAlaLeuAlaLeuGluLeuGlyCysAspThrProLysLeu 345
QY 373 -----CGACAGTAATGCGATGTCACCGTTTCCGATATTCTCGGAA 414
Db 346 LeuGluTyrrIleThrGluArgLeuValGlnProHisTyrAsnGluThrLeuAlaLeuSer 365
QY 415 AATGCCAAG 423
```

```

Db 366 PheSerLys 368
      ::::|
RESULT 4
AD3067
hypotheical protein Atu4154 [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
C:Accession: AD3067
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.;
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; PMID:11743193
A:Accession: AD3067
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-317 <R>
A:Cross-references: GB:AE008689; PIDN:AAL44954.1; PID:g17742609; GSPDB:GN00187
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: Atu4154
A:Map position: linear chromosome

Alignment Scores:
Pred. No.: 1.38 Length: 317
Score: 84.00 Matches: 42
Percent Similarity: 43.70% Conservative: 17
Best Local Similarity: 31.11% Mismatches: 53
Query Match: 10.76% Indels: 23
DB: 2 Gaps: 8

US-09-856-221-1 (1-449) x AD3067 (1-317)
QY 443 ATCAGTTGGTCACT-----CCCCACTTGGCATTTTCCAGATAATCGGAAAA 396
Db 62 ValSerTrpValGluPheThrAlaGlyProLeuValGluAlaLeuAsnValGlySer 81
QY 395 CGGTACATCGGCATTACTGTCGAGGCAAAATCCCTCCG----- 357
Db 82 IleAsnValGlyTrpThr-----GlyAspAlaProPheGlyGlnAlaGly 99
QY 356 CTTGTGAGCGGTGACGACACTAAGCAGTACGACGAGTCTGCTGTCGCATAG 297
Db 100 SerAlaIleValTyValAlaAlaLeuProSer---AsnGlyLysGlyGluAlaIlePhe 118
QY 296 ATGGGTAAATGACACGCTGACCATCAATGTCAGATTATGACGTAAGTTATAGCGGT 237
Db 119 ThrLysProGluSerGlyIleLysSerVal-AlaAspLeuLysGlyLysValGlyVa 138
QY 236 TGTGCTA---ACGTCTGCCAGTAACCTTTCAGATTTTGTATCTGAGGAGGACAAA 180
Db 138 lclYlsglyThrSerAlaHisAsnLeuValAlaAlaLeuGluLysAsnGlyLeuLy 158
QY 179 GCGGTTAACGAATTTTCAG-----TAGCTTCCGCGCAAGCCGCGGTTGACCCAGT 126
Db 158 s---LeuSerAspIleAspValThrTyLeuSerProAla-----AspAlaAl 173
QY 125 TGCCGCATTTTATGCTAATGATCAGCAATGGTTTGACTGG 83
Db 173 aAlaAlaPheAlaSerAspLysIleAspAlaTrpAlaValTrp 187

RESULT 5
F98219
hypotheical protein AGR_L_1399 [imported] - Agrobacterium tumefaciens (strain C58, Cere
C:Species: Agrobacterium tumefaciens
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 11-Jan-2002
C:Accession: F98219
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Ouello, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;

Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium
A:Reference number: A97359; PMID:11743194
A:Accession: F98219
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-338 <R>
A:Cross-references: GB:AE007870; PIDN:AAK89280.1; PID:g15159111; GSPDB:GN00170
C:Genetics:
A:Gene: AGR_L_1399
A:Map position: linear chromosome

Alignment Scores:
Pred. No.: 1.39 Length: 338
Score: 84.00 Matches: 42
Percent Similarity: 43.70% Conservative: 17
Best Local Similarity: 31.11% Mismatches: 53
Query Match: 10.76% Indels: 23
DB: 2 Gaps: 8

US-09-856-221-1 (1-449) x F98219 (1-338)
QY 443 ATCAGTTGGTCACT-----CCCCACTTGGCATTTTCCAGATAATCGGAAAA 396
Db 83 ValSerTrpValGluPheThrAlaGlyProLeuValGluAlaLeuAsnValGlySer 102
QY 395 CGGTACATCGGCATTACTGTCGAGGCAAAATCCCTCCG----- 357
Db 103 IleAsnValGlyTrpThr-----GlyAspAlaProPheGlyGlnAlaGly 120
QY 356 CTTGTGAGCGGTGACGACACTAAGCAGTACGACGAGTCTGCTGTCGCATAG 297
Db 121 SerAlaIleValTyValAlaAlaLeuProSer---AsnGlyLysGlyGluAlaIlePhe 139
QY 296 ATGGGTAAATGACACGCTGACCATCAATGTCAGATTATGACGTAAGTTATAGCGGT 237
Db 140 ThrLysProGluSerGlyIleLysSerVal-AlaAspLeuLysGlyLysValGlyVa 159
QY 236 TGTGCTA---ACGTCTGCCAGTAACCTTTCAGATTTTGTATCTGAGGAGGACAAA 180
Db 159 lclYlsglyThrSerAlaHisAsnLeuValAlaAlaLeuGluLysAsnGlyLeuLy 179
QY 179 GCGGTTAACGAATTTTCAG-----TAGCTTCCGCGCAAGCCGCGGTTGACCCAGT 126
Db 179 s---LeuSerAspIleAspValThrTyLeuSerProAla-----AspAlaAl 194
QY 125 TGCCGCATTTTATGCTAATGATCAGCAATGGTTTGACTGG 83
Db 194 aAlaAlaPheAlaSerAspLysIleAspAlaTrpAlaValTrp 208

RESULT 6
T41977
hypotheical protein U76 - human herpesvirus 7 (strain JI)
C:Species: human herpesvirus 7
A:Variety: strain JI
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000
C:Accession: T41977
R:Nicholas, J.
submitted to the EMBL Data Library, December 1995
A:Description: Determination and analysis of the complete nucleotide sequence of huma
A:Reference number: Z22022
A:Accession: T41977
A:Status: Preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-640 <NIC>
A:Cross-references: EMBL:U43400; PIDN:AAK54737.1
A:Experimental source: strain JI
C:Genetics:
A:Note: U76
C:Superfamily: varicella-zoster virus gene 54 protein

Alignment Scores:
Pred. No.: 6.59 Length: 640

```

QY 341 ACAGGACGCATTAAGCATACGCGAGGTATTGGTGTCTGTCATGATAGATG... :  
DB 129 ProLeuSerLeuPheAsnAspSerAlaPheLeu-HisGluPheIleLeuIl 148  
QY 290 AATGCAACGGCTGACCATCAATTGTGCAGATTATGACGTAAGTTATATAGCGGTGTGCT 231  
DB 148 eargThrAlaGlyAspGluLeuAldGluArgLleLysLeuTyrlauaspValyLe 168  
QY 230 A-----ACGTCGCCAGTAACTTTGCGAGTTTGCT 201  
DB 168 uPhePheLeuTrpLeuIleGlyAsnThrCystThrAsnGlnGlyPheTyrlauPheCy 188  
QY 200 TTTATCTGAGGAAGAACACACGGGTAAACGAAATTTGCAGTACGTTTCGCCGCAAGGCC 141  
DB 188 sLeuAsnGluSerGly-----ArgLeu-----TyrValSerProMetGluTh 202  
QY 140 GCGCGTTGACGAGTTCCGCGCATTTTTATGCTGATAAT 104  
DB 202 rArgSerLeuAlaValGlnGluTrpTyrlauGluGluAsp 214  
  
RESULT 8  
C69681  
peptide synthetase ppSC - Bacillus subtilis  
N;Alternate names: peptide synthetase 3  
C;Species: Bacillus subtilis  
C;Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 03-Nov-2000  
C;Accession: C69681; 140458; S49135  
R;Kunst, F.; Ogasawara, N.; Moser, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber  
C.; Bron, S.; Brouillet, S.; Busch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;  
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,  
Nature 390, 249-256, 1997  
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal  
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hulio, M.  
Koeter, P.; Koningsberg, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino  
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau  
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Poll, T.M.; Portete  
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanl  
A;Authors: Schleich, S.; Schroeter, R.; Scofield, F.; Sekiguchi, J.; Setokawa, A.; Se  
akeuchi, M.; Tanakoshi, A.; Tanaka, T.; Terpsira, P.; Toynon, P.; Uchiya  
T.; Winters, P.; Wibat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida  
A;Authors: Yoshikawa, H.F.; Zamojstain, E.; Yoshikawa, H.; Danchin, A.  
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtili  
A;Reference number: A65580; MUID:98044033; PMID:9384377  
A;Accession: C69681  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-2555 <KUN>  
A;Cross-references: GB:D99113; GB:AL009126; NID:g2634090; PIDN:CAB13715.1; PID:g26342  
A;Experimental source: strain 168  
E;Tognoni, A.; Franchi, E.; Magistrelli, C.; Colombo, E.; Cosmina, P.; Grandi, G.  
Microbiology 141, 645-648, 1995  
A;Title: A putative new peptide synthase operon in Bacillus subtilis: partial charact  
A;Reference number: 140454; MUID:95227362; PMID:7711903  
A;Accession: I40458  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-859 <RES>  
A;Cross-references: EMBL:Z34883; NID:g509465; PID:g509470  
A;Accession: C69681

A:Gene: ppsC; pps3  
 S:Superfamily: carrier protein; phosphopantetheine; phosphoprotein  
 K:Keywords: acetate-CoA ligase homology; acyl carrier protein  
 F:510-953/domain: acetate-CoA ligase homology <ACP>  
 F:510-953/domain: acyl carrier protein homology <ACL1>  
 F:970-1038/domain: acetate-CoA ligase homology <ACL2>  
 F:1551-1990/domain: acetate-CoA ligase homology <ACL2>  
 F:F:2006-2073/domain: acyl carrier protein homology <ACP2>  
 F:F:1002,2038/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Alignment Scores:	12.9	Length:	2555
Pred. No.:	76.00	Matches:	43
Score:	34.86%	Conservative:	18
Percent Similarity:	34.86%	Mismatches:	44
Best Local Similarity:	24.57%		

TPS-09-856-221-1 (1-449) x B69837 (1-307)

401 GGAAACGGTACATCGCATTACTCTCCGAGGCAATCCCCCTCTTGAGCGGTG 342



A:Residues: 1-407 <KUR>  
A:Cross-references: GB:AE005672; PIDN:AAK75434.1; PID:g14972819; GSPDB:GN00164; TIGR:  
A:Experimental source: strain TIGR4  
C:Genetics:  
A:Gene: SPI336

Alignment Scores:  
Pred. No.: 14.7 Length: 407  
Score: 74.50 Matches: 35  
Percent Similarity: 36.8% Conservative: 26  
Best Local Similarity: 22.2% Mismatches: 54  
Query Match: 9.39% Indels: 42  
DB: 2 Gaps: 7

US-09-856-221-1 (1-449) x A95155 (1-407)

QY 66 CCGTTTGAAGAGCCGCCAGTCAACACCATTCGTGATCATTCATTCAGCATAAATGCGGCA 125  
Db 179 PropheargArgGluGlyGlnAlaThrAsnSerGluThrLeuLysAlaLeuGlyAsnLeu 198  
QY 126 ACTGCGTCAACGCGCGCTTGGCGGCAACGTAC----- 161  
Db 199 AsnProSerArgSerGlyMetSerGlyLysValTyrTyrSerGlyGlyLeuAlaProThr 218  
QY 162 -----TGCATATTCGTTAACCGC 179  
Db 219 LeuValArgGlyLysGlyGluGlyPheLysValAlaIleProCysMetThrProAspArg 238  
QY 180 TTGTGCTCTCTCAGATAACAAAGTCTACTGGCAGACGTTAGCAACG 239  
Db 239 Leu-----AspLysArgGlnAsnGlyArgGlyPheLysGluAsnGln-GluPr 254  
QY 240 CCTATATTAATTA-----CGTCAATATCGACATTCGATGTCAG-----CGTT 284  
Db 254 oMetPheThrLeuAsnThrGlnAspArgTyrGlyIleValValGlyAspLeuProTh 274  
QY 285 GTCATTACCC-----ATCTATGCGACACCGACGATCGTCCGTACTGCTTAG 332  
Db 274 rSerPheLysGluThrGlyArgValTyrGlySerGluGly-----LeuSe 289  
QY 333 TGCTCCGTCACCCCTCAACAGCGGAGGGATTGTCCTCGACAGTAAATGCGG---AT 389  
Db 289 rProThrLeuThrMetGlnGlyAspLysIleProLysIleLeuIleProGluPr 309  
QY 390 GTACCGTTTCCGATTATCTGGAAATGCGAAGTGGGAGTGACCCAA 438  
Db 309 orleGlnPheLysValArgGluAlaThrLysGlyTyrAlaGln 325

RESULT 13

T31093  
probable protein-tyrosine-phosphatase (EC 3.1.3.48) - medicinal leech  
N:Alternate names: receptor tyrosine phosphatase  
C:Species: Hirudo medicinalis (medicinal leech)  
C>Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 31-Jan-2000  
C:Accession: T31093  
R:Gershon, T.R.; Baker, M.; Nitabach, M.; Wu, P.; Macagno, E.R.  
submitted to the EMBL Data Library, December 1997  
A:Description: Two receptor tyrosine phosphatases expressed by neurons and muscle cell  
A:Reference number: Z20976  
A:Accession: T31093  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1437 <GER>  
A:Cross-references: EMBL:AF017084; NID:g2695656; PID:g2695657; PIDN:AA91461.1  
C:Genetics:  
A:Gene: LAR1  
C:Superfamily: leukocyte antigen-related protein; fibronectin type III repeat homolog  
og  
C:Keywords: glycoprotein; phosphoprotein; phosphoric monoester hydrolase; transmembra

Alignment Scores:  
Pred. No.: 17.3 Length: 1437  
Score: 74.50 Matches: 37

QY 331 AGTGCTGCGTCACCGCTCAAGCGGAGGAGGATTGCTCCGCG 375  
Db 127 ThrProAlaThrProAlaSerGluSerSerGlyIleValProGln 141

RESULT 11

F96660  
protein F2K11.10 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C:Accession: F96660  
R:Neologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Hughes, B.; Huizar, L.  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziani,  
Pizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: AB6141; MUID:21016719; PMID:11130712  
A:Accession: F96660  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-635 <STO>  
A:Cross-references: GB:AE005173; NID:g6633854; PIDN:AAF19713.1; GSPDB:GN00141  
C:Genetics:  
A:Gene: F2K11.10  
A:Map position: 1

Alignment Scores:  
Pred. No.: 13.8 Length: 635  
Score: 75.00 Matches: 20  
Percent Similarity: 46.0% Conservative: 9  
Best Local Similarity: 31.7% Mismatches: 28  
Query Match: 9.60% Indels: 6  
DB: 2 Gaps: 1

US-09-856-221-1 (1-449) x F96660 (1-635)

QY 449 AACTGTATCAGTTGGTGC-----ACTCCCGACITGGCATTTCCACA 408  
Db 12 AsnCysTyrGlnPheThrAsnProAspLeuLeuAsnThrProGluSerGluGlnSerAsn 31  
QY 407 ATATCGGAACGATACATCGCATTTACTGTCCGAGGCAATCCCTCCGCGCTTGTGAG 348  
Db 32 VallleGlySerIleThrSerValProValAsnAspGlyProValProProLeuGlu 51  
QY 347 GCGGTGACGAGCACTAGCAGTACGACGAGTCTGCTGTCGATAGATGGTAAT 288  
Db 52 LeuAspSerAlaAlaValSerThrSerThrSerProValGlnAlaLeuGlyHis 71  
QY 287 GACACGCGC 279  
Db 72 AspSerGly 74

RESULT 12

A95155  
hypothetical protein SPI336 [imported] - Streptococcus pneumoniae (strain TIGR4)  
C:Species: Streptococcus pneumoniae  
C>Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 03-Aug-2001  
C:Accession: A95155  
R:Rettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid  
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapfel,  
nson, T.; Hickey, E.K.; Holt, I.E.  
Science 293, 498-506, 2001  
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,  
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.  
A:Reference number: A95000; MUID:21357209; PMID:11463916  
A:Accession: A95155  
A>Status: preliminary  
A:Molecule type: DNA





QY 313 GATCCGTCGTTACTGCTTAGTGTCCGTCACCGCCTCACAGCGCGAGGGGATTTGCCT 372  
Db 102 ThrProMetThrProIleSerProAlaThrProAlaSerGluSerSerGlyIleValPro 121  
QY 373 CGG 375  
Db 122 Gln 122

Search completed: January 15, 2003, 15:56:15  
Job time : 26.6921 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model  
Run on: January 15, 2003, 15:45:23 ; Search time 10.6966 Seconds  
(without alignments)  
1826.456 Million cell updates/sec

Title: US-09-856-221-4  
Perfect score: 605  
Sequence: 1 gtagagcgagctattgca.....atttggaagatgaacgg 332

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delep 6.0, Delext 7.0

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+n2p.model -DEV=rlp  
-Q=/cgn2\_6/ptodata/1/iaa/5A.COMB.pep.\*  
-DB=issued\_patents\_AA -OFF=fastan -SUFFIX=rai -MINMATCH=0.1 -IOOFCI=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09856221.ecgn.1.1.24.@runat.15012003.153926.12521 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=10 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued\_Patents\_AA.\*  
1: /cgn2\_6/ptodata/1/iaa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/5B.COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/6A.COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/6B.COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/6C.COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/6D.COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	440	72.7	2522	US-09-251-645-13	Sequence 13, Appl
2	70.5	11.7	507	US-08-363-475-19	Sequence 19, Appl
3	70.5	11.7	532	US-08-363-475-22	Sequence 22, Appl
4	67.5	11.2	680	US-08-674-351-2	Sequence 2, Appl
5	67.5	11.2	707	US-09-021-560-4	Sequence 4, Appl
6	67.5	11.2	744	US-09-021-560-2	Sequence 2, Appl
7	64	10.6	218	US-09-081-689-6	Sequence 6, Appl
8	64	10.6	218	US-09-303-984-16	Sequence 16, Appl
9	64	10.6	218	US-09-073-541A-16	Sequence 16, Appl
10	64	10.6	464	US-09-604-978-7	Sequence 7, Appl
11	64	10.6	919	US-08-377-503-2	Sequence 2, Appl
12	64	10.6	919	US-08-178-019-2	Sequence 2, Appl

13	63	10.4	469	4	US-09-131-750-29	Sequence 29, Appl
14	61.5	10.2	154	4	US-09-404-670-4	Sequence 4, Appl
15	60	9.9	921	1	US-07-718-575-14	Sequence 14, Appl
16	60	9.9	921	1	US-08-481-206-14	Sequence 14, Appl
17	60	9.9	921	2	US-08-486-269A-14	Sequence 14, Appl
18	59.5	9.8	154	4	US-09-134-001C-2992	Sequence 2992, Ap
19	59.5	9.8	503	4	US-09-562-737-63	Sequence 63, Appl
20	58.5	9.7	430	4	US-09-182-816-28	Sequence 28, Appl
21	58.5	9.7	430	4	US-09-471-528-28	Sequence 28, Appl
22	58.5	9.7	430	4	US-09-634-530-28	Sequence 28, Appl
23	58.5	9.7	456	4	US-09-134-001C-3771	Sequence 3771, Ap
24	58.5	9.7	465	4	US-09-182-816-23	Sequence 23, Appl
25	58.5	9.7	465	4	US-09-471-528-23	Sequence 23, Appl
26	58.5	9.7	465	4	US-09-634-530-23	Sequence 23, Appl
27	58	9.6	484	2	US-08-252-493C-9	Sequence 9, Appl
28	58	9.6	484	3	US-09-276-197-9	Sequence 9, Appl
29	58	9.6	497	4	US-09-457-046B-65	Sequence 65, Appl
30	58	9.9	1165	1	US-08-240-357-2	Sequence 2, Appl
31	58	9.6	2233	2	US-08-569-853-1	Sequence 1, Appl
32	58	9.6	2233	3	US-08-987-439-1	Sequence 1, Appl
33	57.5	9.5	286	4	US-09-172-952-17	Sequence 17, Appl
34	57.5	9.8	440	4	US-08-759-628-4	Sequence 4, Appl
35	57.5	9.5	907	3	US-08-863-102-2	Sequence 2, Appl
36	57	9.4	222	4	US-09-134-001C-4748	Sequence 4748, Ap
37	57	9.4	741	4	US-09-001-984C-106	Sequence 106, App
38	56.5	9.3	119	1	US-08-340-539A-16	Sequence 16, Appl
39	56.5	9.3	342	4	US-08-096-181A-12	Sequence 12, Appl
40	56.5	9.3	342	4	US-08-096-181A-14	Sequence 14, Appl
41	56.5	9.3	342	5	PCT-US94-08326-12	Sequence 12, Appl
42	56.5	9.3	342	5	PCT-US94-08326-14	Sequence 14, Appl
43	56.5	9.3	361	4	US-08-096-181A-8	Sequence 8, Appl
44	56.5	9.3	361	5	PCT-US94-08326-8	Sequence 8, Appl
45	56.5	9.3	363	4	US-08-096-181A-10	Sequence 10, Appl

ALIGNMENTS

RESULT 1  
US-09-251-645-13  
; Sequence 13, Application US/09251645  
; Patent No. 6281413  
; GENERAL INFORMATION:  
; APPLICANT: Kramer, Vance.C.  
; APPLICANT: Morgan, Michael K.  
; APPLICANT: Anderson, Arne R.  
; APPLICANT: Hart, Hope  
; APPLICANT: Warren, Gregory W.  
; APPLICANT: Dunn, Martha  
; APPLICANT: Chen, Jeng S.  
; TITLE OF INVENTION: NOVEL INSECTICIDAL TOXINS FROM PHOTOPHABDUS LUMINESCENS  
; TITLE OF INVENTION: AND NUCLEIC ACID SEQUENCES CODING THEREFOR  
; FILE REFERENCE: CCI963/A  
; CURRENT APPLICATION NUMBER: US/09/251,645  
; CURRENT FILING DATE: 1999-02-17  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 13  
; LENGTH: 2522  
; TYPE: PRT  
; ORGANISM: Photorhabdus luminescens  
US-09-251-645-13

Alignment Scores:  
Pred. No.: 8.2e-53  
Score: 440.00  
Length: 2522  
Matches: 84  
Percent Similarity: 88.18%  
Conservative: 13  
Best Local Similarity: 76.36%  
Mismatch: 13  
Query Match: 72.73%  
Indels: 0  
DB: Gaps: 0

US-09-856-221-4 (1-332) x US-09-251-645-13 (1-2522)

QY 3 GAAGCGGAGTATTGCAAAACTATCTGGAACCAACGACAGCAAACTCAGGCACAG 62

Interference?

|||||  
Db 2224 GluAlaValLeuGlnLysThrSerLeuLysThrGlnGlnGlnThrGlnAlaGln 2243  
QY 63 CTGGCTTCTCAAGCAAAATTCAGCAATACAGCGTGTGTATAACTGGCTAGCTGGCGGA 122  
|||||  
Db 2244 LeuAlaPheLeuGlnArgLysPheSerAsnGlnAlaLeuTyraAsnTrpLeuArgGlyArg 2263  
QY 123 TTGGCGGCTATTATATACAGTTTATGACTTGGCTGCTTTCCTGCTGTTGATGCTGAA 182  
|||||  
Db 2264 LeuAlaAlaIleTyPheGlnPheTyraPheLeuAlaIleAlaArgCysLeuMetAlaGlu 2283  
QY 183 CAACCTACAGTATGAATTAACGATAAAGCTGTACGCTTCATTAAAGCCCGCTGCCGCG 242  
|||||  
Db 2284 GlnAlaTyraTrpGlnIleSerAspSerAlaArgPheIleLysProGlyAlaTrp 2303  
QY 243 CATGGCACTATGCTGGTTGTAGCAGGTGAACCTGTGATGCTGAATTTGGCAGATG 302  
|||||  
Db 2304 GlnGlyThrTyraLacGlyLeuLeuAlaGlyGluThrLeuMetLeuSerLeuAlaGlnMet 2323  
QY 303 GAAAAAATATTGGAAGATGAACGG 332  
Db 2324 GluAspAlaHisLeuArgAspLysArg 2333

RESULT 2  
US-08-363-475-19  
; Sequence 19, Application US/08363475  
; Patent No. 5516679  
; GENERAL INFORMATION:  
; APPLICANT: Chiang, Shu-Jen  
; APPLICANT: Burnett Jr., William V.  
; APPLICANT: Tonzi, Sean M.  
; TITLE OF INVENTION: PENICILLIN V AMIDOHYDROLASE GENE FROM  
; TITLE OF INVENTION: FUSARIUM OXYSPORUM  
; NUMBER OF SEQUENCES: 30  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Thomas R. Savitsky  
; STREET: P.O. Box 4000  
; CITY: Princeton  
; STATE: New Jersey  
; COUNTRY: U.S.A.  
; ZIP: 08543  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/363,475  
; FILING DATE: 23-DEC-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Savitsky, Thomas R.  
; REGISTRATION NUMBER: 31,661  
; REFERENCE/DOCKET NUMBER: ON-0134  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (609) 252 4956  
; TELEFAX: (609) 252 4526  
; INFORMATION FOR SEQ ID NO: 19:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 507 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-363-475-19  
Alignment Scores:  
Pred. No.: 0.33  
Score: 70.50  
Percent Similarity: 37.25%  
Best Local Similarity: 20.26%  
Query Match: 11.65%  
DB: 1  
Length: 507  
Matches: 31  
Conservative: 26  
Mismatch: 45  
Indels: 51  
Gaps: 5

US-09-856-221-4 (1-332) x US-08-363-475-19 (1-507)  
QY 3 GAAGCGCAGTATTGCAAAACCACTATCTGGAACCAACAGGACCAAACTCAGGCACAG 62  
|||||  
Db 259 GluAlaLeuIleCysGlyLysGly-----GlnThrGluAsnCysLeuThrLysAlaLys 276  
QY 63 CTGGCTTCTCAACAAACAAATTCAGC----- 89  
Db 277 IleGluThrValArgLysValPheSerProLeuTyThrThrAsnGluThrTyValTy 296  
QY 90 -----AATACACGGTGTATAACTGCTACGCTACGCGGCGGCTTGGCGGCT 131  
Db 297 ProArgAlaValProGlyAlaAsnAlaLeuPheAsnPheValAlaGluThrProPhe 316  
QY 132 ATTATAT-----TATCAGTTT----- 146  
Db 317 ValTySerThrGluTrpTyThrGlnTyValIleTrpGluAspProGluTrpAsnProAsp 336  
QY 147 -----TATGACTTGGCTGTTTCCCTGCTGTTTGCATGGCTGCAAACT 188  
Db 337 ThrIleGlyProLysAspTyraPargGlyAlaGluMetAsnProTyraPargIleGluThr 356  
QY 189 TACAGTATGAATTAACGATAAAGCTGTACGCTTCATTAAAGCCCGGCTGCGCATGGC 248  
Db 357 TrpGluGlyAspLeuSerLysPheArgLysArgGlyAsnLysMetIleHisTrpHisGly 376  
QY 249 ACTTATGCTGTTTGTAGCAGGTGAA----- 275  
Db 377 LeuGlnAspGlyLeuIleSerAlaGluAsnSerAspAspTyThrAsnHisValSerArg 396  
QY 276 ACCTGTGATGCTGAATTTGGCAGCAGATGGGAAAAAACTAT 314  
Db 397 ThrMetGlyLeuAsnSerSerGlnLeuAspGlnPheTy 409  
RESULT 3  
US-08-363-475-22  
; Sequence 22, Application US/08363475  
; Patent No. 5516679  
; GENERAL INFORMATION:  
; APPLICANT: Chiang, Shu-Jen  
; APPLICANT: Burnett Jr., William V.  
; APPLICANT: Tonzi, Sean M.  
; TITLE OF INVENTION: PENICILLIN V AMIDOHYDROLASE GENE FROM  
; TITLE OF INVENTION: FUSARIUM OXYSPORUM  
; NUMBER OF SEQUENCES: 30  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Thomas R. Savitsky  
; STREET: P.O. Box 4000  
; CITY: Princeton  
; STATE: New Jersey  
; COUNTRY: U.S.A.  
; ZIP: 08543  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/363,475  
; FILING DATE: 23-DEC-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Savitsky, Thomas R.  
; REGISTRATION NUMBER: 31,661  
; REFERENCE/DOCKET NUMBER: ON-0134  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (609) 252 4956  
; TELEFAX: (609) 252 4526  
; INFORMATION FOR SEQ ID NO: 22:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 532 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear



```

: NAME: MURPHY JR, GERALD M
: REGISTRATION NUMBER: 28,977
: REFERENCE/DOCKET NUMBER: 825-144P
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703) 205-8000
: TELEFAX: (703) 205-8050
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 744 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-09-021-560-2

```

Alignment Scores:		
Pred. NO.:	1.01	744
Score:	67.50	27
% Similarity:	35.79%	7
Best Local Similarity:	28.4%	Conservative:
Query Match:	11.1%	Mismatches:
DB:	4	Indels:
		Gaps:

US-09-856-221-4 (1-332) x US-09-021-560-2 (1-744)

Qy	66	GC	TTCTCC	TACAAAGCAAA	TTTCAGCATACACAGCCTTGCTATAC	TGCTACGTGGCGG	CAATG	125
Db	599	Ala	Pha	Ile	Le	Ser	Ser	618
Qy	126	GGG	GCTATT	TATATAC	AGTTTATGACTTGGCTGTTTCCCTGCTGTTT	TGATGGCT	CAACAA	185
Db	619	Ala	Asp	Ala	Leu	Tyr	Asn	625
Qy	186	ACT	TACCA	GATGAT	GAATGAACAGTAAGCTGTACGCTTCATTAAGCCGGTGCCTGGC	AT	245	
Db	626	---	---	---	Ile	Asn	Asp	638
Qy	246	GGC	ACTTAT	GCTGTTGTTT	TAGCAGGPGAAACCTTG	---	---	281
Db	639	Lys	Leu	Ser	Val	Gly	Leu	658
Qy	282	---	---	---	ATG	CGAATTTGGC	CAGATGGAAAAA	323
Db	659	Glu	Tyr	Val	Iasn	Leu	Ala	673

RESULT 7  
US-09-081-689-6  
; Sequence 6, Application US/09081689  
: Patent No. 6165992

```

1  GENERAL INFORMATION:
2
3  APPLICANT: Wallis, Nicola G.
4
5  APPLICANT: Zalacain, Madgalenaa
6
7  APPLICANT: Throup, John
8
9  APPLICANT: Biswas, Sanjoy
10
11 TITLE OF INVENTION: Histidine Kinase
12
13 NUMBER OF SEQUENCES: 9
14
15 CORRESPONDENCE ADDRESS:
16
17 ADDRESS: Dechert, Price & Rhoads
18
19 STREET: 4000 Bell Atlantic Tower, 1717 Arc
20
21 CITY: Philadelphia
22
23 STATE: PA
24
25 COUNTRY: USA
26
27 ZIP: 19103-2793
28
29 COMPUTER READABLE FORM:
30
31 MEDIUM TYPE: Diskette
32
33 COMPUTER: IBM Compatible
34
35 OPERATING SYSTEM: Windows 95
36
37 SOFTWARE: FASTSEQ for Windows Version 2.0b
38
39 CURRENT APPLICATION DATA:
40
41 APPLICATION NUMBER: US/09/081.689
42
43 FILING DATE:
44
45 CLASSIFICATION:
46
47 PRIOR APPLICATION DATA:
48
49 APPLICATION NUMBER: 60/048.347

```



10



QY 66 GCGCTTCTACAAAGCAAAATTCAGCAATACAGCGTTGTATTAACCTAGCTGGCGGATTTG 125  
Db 313 GlyLeuLeuAspGlyValMetMetThrAspAlaAlaLeuLeuTyrAspAlaValHisIle 332  
QY 126 GCGGCTATTTATATACAGTTTATGACTTGGCTGTTCCCTGTTGATGGCTGAACAA 185  
Db 333 ValSerValCysTyrGlnArgAlaProGlnMetThrValAsnSerLeuGlnCysHisGln 352  
QY 186 ACTTACCAGTATGAATGAACGATAAAGCTGTACGCTTCATTAAAGCCGCTGGCGCAT 245  
Db 353 HisLysAlaTyrArgPheGlyGlyArgPheMetAsnPheIleLysGluAlaGlnTyrGlu 372  
QY 246 GCGCTATTCGCTGTTGTTAGCAGGTGAACAC-----TTGATGCTG 287  
Db 373 GlyLeuThrGlyArgIleValPheAsnLysThrSerGlyLeuArgThrAspPheAspLeu 392  
QY 288 AATTGGCACAGATGAAAAAACTATTGGAAAAA 323  
Db 393 AspilelleSerLeuLysGluAspGlyLeuGluLys 404

RESULT 12  
US-08-178-019-2  
; Sequence 2, Application US/08178019  
; Patent No. 633161  
; GENERAL INFORMATION:  
; APPLICANT: Kamboj, Rajender  
; APPLICANT: Elliott, Candace E.  
; APPLICANT: Nutt, Stephen L.  
; TITLE OF INVENTION: Kainate-binding, Human CNS Receptors of  
; TITLE OF INVENTION: The EA5 Family  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: Suite 500, 3000 K Street  
; CITY: N.W.  
; STATE: Washington, D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/178.019  
; FILING DATE: 06-JAN-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bent, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 16777/227  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 672-5300  
; TELEFAX: (202) 672-5399  
; TELEX: 904136

INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 919 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-178-019-2

Alignment Scores:  
Pred. No.: 3,41 Length: 919  
Score: 64.00 Matches: 22  
Percent Similarity: 40.18% Conservative: 23  
Best Local Similarity: 19.64% Mismatches: 61  
Query Match: 10.58% Indels: 6  
DB: 4 Gaps: 1

US-09-856-221-4 (1-332) x US-08-178-019-2 (1-919)

QY 6 GCGCAGTATTGCAAAAAAATCTCTGGAAACCCAAACAGGCACAACTCAGGCACAGCTG 65  
Db 293 SerAlaIleValGluLysTrpSerMetGluArgLeuGlnAlaAlaProArgAlaGluSer 312  
QY 66 GCGTTCCTACAAAGCAAAATTCAGCAATACAGCGTTGTATTAACCTAGCTGGCGGATTTG 125  
Db 313 GlyLeuLeuAspGlyValMetMetThrAspAlaAlaLeuLeuTyrAspAlaValHisIle 332  
QY 126 GCGGCTATTTATATACAGTTTATGACTTGGCTGTTCCCTGTTGATGGCTGAACAA 185  
Db 333 ValSerValCysTyrGlnArgAlaProGlnMetThrValAsnSerLeuGlnCysHisGln 352  
QY 186 ACTTACCAGTATGAATGAACGATAAAGCTGTACGCTTCATTAAAGCCGCTGGCGCAT 245  
Db 353 HisLysAlaTyrArgPheGlyGlyArgPheMetAsnPheIleLysGluAlaGlnTyrGlu 372  
QY 246 GCGCTATTCGCTGTTGTTAGCAGGTGAACAC-----TTGATGCTG 287  
Db 373 GlyLeuThrGlyArgIleValPheAsnLysThrSerGlyLeuArgThrAspPheAspLeu 392  
QY 288 AATTGGCACAGATGAAAAAACTATTGGAAAAA 323  
Db 393 AspilelleSerLeuLysGluAspGlyLeuGluLys 404

RESULT 13  
US-09-131-750-29  
; Sequence 29, Application US/09131750A  
; Patent No. 6361564  
; GENERAL INFORMATION:  
; APPLICANT: Kaiser, Chris A.  
; APPLICANT: Frand, Allison R.  
; TITLE OF INVENTION: EUKARYOTIC DISULFIDE BOND-FORMING  
; TITLE OF INVENTION: PROTEINS AND RELATED MOLECULES AND METHODS  
; FILE REFERENCE: 01997/502002  
; CURRENT APPLICATION NUMBER: US/09/131,750A  
; CURRENT FILING DATE: 1998-08-10  
; EARLIER FILING DATE: 1997-08-12  
; NUMBER OF SEQ ID NOS: 29  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 29  
; TYPE: PRT  
; ORGANISM: Homo sapiens, Mus musculus, Rattus norvegicus  
US-09-131-750-29

Alignment Scores:  
Pred. No.: 3,62 Length: 469  
Score: 63.00 Matches: 24  
Percent Similarity: 38.05% Conservative: 19  
Best Local Similarity: 21.24% Mismatches: 30  
Query Match: 10.41% Indels: 40  
DB: 4 Gaps: 3

US-09-856-221-4 (1-332) x US-09-131-750-29 (1-469)

QY 12 GTATTGCAAAAAAATCTATCTGGAACCAACAGGCACAACTCAGGCACAGCTGGCGCTTC 71  
Db 17 LeuLeuGlnSerGlyHisGlyGluGlnArgProGluThrAlaAlaGlnArgCysPhe 36  
QY 72 CTACAA-----ACCAAAATTC 86  
Db 37 CysGlnValSerGlyTyrLeuAspCysThrCysAspValGluThrIleAspArgPhe 56  
QY 87 AGCAATACAGCGTTGTATTAACCTAGCTGGCGGATTTGCGGCTATTTATATCATGTTT 146  
Db 57 AsnAsnTyrArgLeuPheProArgLeuGlnLysLeuLeuGluSerAspTyrPheArgTyr 76  
QY 147 TAT----- 149  
Db 77 TyrLysValAsnLeuLysArgProCysProIleTrpAsnAspIleSerGlnCysGlyArg 96



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model  
Run on: January 15, 2003, 15:46:43 ; Search time 6.22838 Seconds  
(without alignments)  
2119.222 Million cell updates/sec

Title: US-09-856-221-4  
Perfect score: 605  
Sequence: 1 gtagagcggcagtgattgcaaa.....atttgaaaaagatgaacgg 332

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 120991 seqs, 19878514 residues

Total number of hits satisfying chosen parameters: 241982

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp  
-Q/cgn2\_1/USPTO\_spool/US09856221/runat\_15012003\_153926\_12555/app.query.fasta\_1.1948  
-DB=Published.Applications\_AA -QEMT=fastan -SUFFIX=rapb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -STAR=1 -END=1 -MARIX=blomsm62  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=plc -NORM=ext -HEADSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US09856221@cgn\_1\_14\_@runat\_15012003\_153926\_12555  
-NCPU=6 -ICPU=3 -NO\_XLPXY -NO\_JMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications\_AA:  
1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	446	73.7	2516	10	US-09-817-514A-2 Sequence 2, Appli
2	418	69.1	2504	10	US-09-817-514A-8 Sequence 8, Appli
3	68	11.2	696	10	US-09-815-242-5443 Sequence 5443, Ap
4	68	11.2	698	10	US-09-815-242-12325 Sequence 12325, A

65	11.1	312	10	US-09-886-055-171	Sequence 171, App
64	10.6	218	10	US-09-737-068-6	Sequence 6, Appli
64	10.6	698	9	US-09-801-220-4	Sequence 4, Appli
63.5	10.5	708	8	US-08-834-666A-2	Sequence 2, Appli
62.5	10.3	489	10	US-09-815-242-13932	Sequence 13932, A
62.5	10.3	733	8	US-08-834-666A-4	Sequence 4, Appli
62.5	10.3	745	8	US-08-834-666A-6	Sequence 6, Appli
61	10.1	47	10	US-09-864-761-37729	Sequence 37729, A
61	10.4	359	10	US-09-764-864-996	Sequence 996, App
61	10.1	487	10	US-09-815-242-12036	Sequence 12036, A
60.5	10.0	163	10	US-09-861-451A-80	Sequence 80, Appl
60	9.9	323	10	US-09-815-242-10578	Sequence 10578, A
60	9.9	406	9	US-09-738-626-4686	Sequence 4686, Ap
59.5	9.8	777	10	US-09-864-761-37995	Sequence 37995, A
59	10.1	159	10	US-09-864-761-46740	Sequence 46740, A
59	10.1	207	9	US-09-811-284-168	Sequence 168, App
58.5	9.7	165	10	US-09-864-761-34905	Sequence 34905, A
58.5	9.7	489	10	US-09-815-242-10215	Sequence 10215, A
58	9.9	124	10	US-09-811-284-165	Sequence 165, App
58	9.6	497	9	US-09-866-570A-65	Sequence 65, Appl
58	9.6	497	10	US-09-866-572A-65	Sequence 65, Appl
58	9.6	574	10	US-09-953-688A-1	Sequence 1, Appli
57.5	9.5	368	10	US-10-025-222A-40	Sequence 40, Appl
57.5	9.5	426	10	US-09-801-368-308	Sequence 308, App
57.5	9.8	442	9	US-09-764-853-667	Sequence 667, App
57.5	9.5	716	9	US-09-870-759-55	Sequence 55, Appl
57	9.7	261	10	US-10-008-355-7	Sequence 7, Appli
56.5	9.6	419	9	US-09-764-864-961	Sequence 961, App
56.5	9.3	429	10	US-09-738-626-4134	Sequence 4134, Ap
56.5	9.3	430	10	US-09-815-242-5541	Sequence 5541, Ap
56.5	9.3	430	10	US-09-815-242-12226	Sequence 12226, A
56.5	9.3	430	10	US-09-815-242-12825	Sequence 12825, A
56.5	9.3	430	10	US-09-815-242-13050	Sequence 13050, A
56.5	9.3	683	9	US-09-738-626-6961	Sequence 6961, Ap
56	9.3	235	10	US-09-799-777-67	Sequence 67, Appl
56	9.3	265	10	US-09-815-242-10621	Sequence 10621, A
56	9.5	270	10	US-09-841-132-404	Sequence 404, App
56	9.3	291	10	US-09-829-631A-10	Sequence 10, Appl
56	9.5	394	9	US-09-738-626-4273	Sequence 4273, Ap
56	9.3	437	10	US-09-829-631A-8	Sequence 8, Appli

ALIGNMENTS

RESULT 1  
US-09-817-514A-2  
; Sequence 2, Application US/09817514A  
; Patent No. US20020078478A1  
; GENERAL INFORMATION:  
; APPLICANT: french-Constant, Richard  
; APPLICANT: Bowen, David  
; APPLICANT: Rocheleau, Thomas  
; APPLICANT: Waterfield, Nicholas  
; TITLE OF INVENTION: DNA SEQUENCES FROM PHOTORHABDUS LUMINESCENS  
; FILE REFERENCE: 61645  
; CURRENT APPLICATION NUMBER: US/09/817,514A  
; PRIOR FILING DATE: 2000-03-26  
; PRIOR APPLICATION NUMBER: US 60/191806  
; NUMBER OF SEQ IDS NOS: 8  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 2  
; LENGTH: 2516  
; TYPE: PRT  
; ORGANISM: Photorhabdus luminescens  
US-09-817-514A-2  
==2002007478

Alignment Scores:  
Pred. No.: 5.98e-49 Length: 2516  
Score: 446.00 Matches: 86  
Percent Similarity: 89.09% Conservative: 12  
Best Local Similarity: 78.18% Mismatches: 12

*Interference*

```
Query Match: 73.72% Indels: 0
DB: 10 Gaps: 0
US-09-856-221-4 (1-332) x US-09-817-514A-2 (1-2516)
QY 3 GAAGCGCAGTATTGCAAAAACATCTGGAACCCACAGGCACAACTCAGGCACAG 62
DB 2218 GluAlaIaValLeuGlnLysThrSerLeuLysThrGlnGlnGlnThrGlnSerGln 2237
QY 63 CTGGCGCTTCTCAACAAGCAAAATCAGCAATACAGCGTTGTATTAACCTAGCGTACGGGGCA 122
DB 2238 LeuAlaPheLeuGlnArgLysPheSerAsnGlnAlaLeuTyrAsnTrpLeuArgGlyArg 2257
QY 123 TTGGCGGCTATTATTATACGCTTTATGACTTGCGTGTTCCTGCTGTGTTGATGCTGAA 182
DB 2258 LeuAlaIaIleTyrPheGlnPheTyrAspLeuAlaValAlaArgCysLeuMetAlaGln 2277
QY 183 CAAACTTACCATGATGAATTGAACGATAAAGCTGTACGCTTCATTAAAGCCCGGTGCTGG 242
DB 2278 GlnAlaTyrArgTrpGluLeuAsnAspSerAlaArgPheIleLysProGlyAlaTrp 2297
QY 243 CATGGCACTTATGCTGGTTTGTAGCAGGTGAAACCTTGTATGCTGAATTTGGCAGATG 302
DB 2298 GlnGlyThrTyrAlaGlyLeuAlaGlyGluThrLeuMetLeuSerLeuAlaGlnMet 2317
QY 303 GAAAAAACTATTGTGAAAAAGATGAACGG 332
DB 2318 GluAspAlaHisLeuLysArgAspLysArg 2327
RESULT 2
US-09-817-514A-8
; Sequence 8, Application US/09817514A
; Patent No. US20020078478A1
; GENERAL INFORMATION:
; APPLICANT: ffrench-Constant, Richard
; APPLICANT: Bowen, David
; APPLICANT: Rocheleau, Thomas
; APPLICANT: Waterfield, Nicholas
; TITLE OF INVENTION: DNA SEQUENCES FROM PHOTORHABDUS LUMINESCENS
; FILE REFERENCE: 61645
; CURRENT APPLICATION NUMBER: US/09/817,514A
; CURRENT FILING DATE: 2000-03-26
; PRIOR APPLICATION NUMBER: US 60/191806
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
; LENGTH: 2504
; TYPE: PRT
; ORGANISM: Photorhabdus luminescens
US-09-817-514A-8
Alignment Scores:
Pred. No.: 2,37e-45 Length: 2504
Score: 418.00 Matches: 78
Percent Similarity: 84.55% Conservative: 15
Best local Similarity: 70.91% Mismatches: 17
Query Match: 69.09% Indels: 0
DB: 10 Gaps: 0
US-09-856-221-4 (1-332) x US-09-817-514A-8 (1-2504)
QY 3 GAAGCGCAGTATTGCAAAAACATCTGGAACCCACAGGCACAACTCAGGCACAG 62
DB 2212 GluAlaAlaGluMetGlnLysGlnTyrLeuLysThrGlnGlnAlaGlnAlaGln 2231
QY 63 CTGGCGCTTCTCAACAAGCAAAATCAGCAATACAGCGTTGTATTAACCTAGCGTACGGGGCA 122
DB 2232 LeuThrPheLeuArgSerLysPheSerAsnGlnAlaLeuTyrSerTrpLeuArgGlyArg 2251
QY 123 TTGGCGGCTATTATTATACGCTTTATGACTTGCGTGTTCCTGCTGTGTTGATGCTGAA 182
DB 2252 LeuSerGlyIleTyrPheTyrAspLeuAlaValSerArgCysLeuMetAlaGln 2271
QY 183 CAAACTTACCATGATGAATTGAACGATAAAGCTGTACGCTTCATTAAAGCCCGGTGCTGG 242
DB 627 GlnThrTyrGln-----AlaThrValLysArgIleGluLys----- 638
QY 243 CATGGCACTTATGCTGGTTTGTAGCAGGTGAAACCTTGTATGCTGAATTTGGCAGATG 302
DB 639 TyrGlyAlaPheValGlyLeuPheProGlyLysAspAlaLeuLeuHisIleSerGlnIle 658
QY 303 GAAAAAACTATTGTGAAAAAGATGA 329
DB: 10 Gaps: 2
US-09-815-242-5443
; Sequence 5443, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5443
; LENGTH: 696
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5443
Alignment Scores:
Pred. No.: 1.6 Length: 696
Score: 68.00 Matches: 17
Percent Similarity: 57.14% Conservative: 11
Best local Similarity: 34.69% Mismatches: 13
Query Match: 11.24% Indels: 8
DB: 10 Gaps: 2
US-09-856-221-4 (1-332) x US-09-815-242-5443 (1-696)
QY 183 CAAACTTACCATGATGAATTGAACGATAAAGCTGTACGCTTCATTAAAGCCCGGTGCTGG 242
DB 627 GlnThrTyrGln-----AlaThrValLysArgIleGluLys----- 638
QY 243 CATGGCACTTATGCTGGTTTGTAGCAGGTGAAACCTTGTATGCTGAATTTGGCAGATG 302
DB 639 TyrGlyAlaPheValGlyLeuPheProGlyLysAspAlaLeuLeuHisIleSerGlnIle 658
QY 303 GAAAAAACTATTGTGAAAAAGATGA 329
DB: 10 Gaps: 2
```

Db 659 SerLysAsnArgIleGluLysValGlu 667

RESULT 4

US-09-815-242-12325

Sequence 12325, Application US/09815242

Patent No. US20020061569A1

GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert

APPLICANT: Ohlsen, Kari L.

APPLICANT: Zyskind, Judith W.

APPLICANT: Wall, Daniel

APPLICANT: Carr, Grant J.

APPLICANT: Yamamoto, Robert T.

APPLICANT: Xu, H. Howard

TITLE OF INVENTION: Identification of Essential Genes In

FILE REFERENCE: ELITRA.011A

CURRENT FILING DATE: 2001-03-21

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

NUMBER OF SEQ ID NOS: 14110

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 12325

LENGTH: 698

TYPE: PRT

ORGANISM: Staphylococcus aureus

US-09-815-242-12325

Alignment Scores:

Pred. No.: 1.6 Length: 698

Score: 68.00 Matches: 17

Percent Similarity: 57.14% Conservative: 11

Best Local Similarity: 34.69% Mismatches: 13

Query Match: 11.24% Indels: 8

DB: 10 Gaps: 2

US-09-856-221-4 (1-332) x US-09-815-242-12325 (1-698)

QY 183 CAAACTTACCAGTGAATTAAGCATTAAGCTGTACGCTTCAATTAAGCCCGTGCCTGG 242

Db 627 GlnThrTyrGln-----AlaThrValLysArgIleGluLys----- 638

QY 243 CATGCCACTTATCTGTTGTAGCAGTGAACCTTGATGCTGAATTTGCCACAGATG 302

Db 639 TgryGlyAlaPheValGluPheProGlyLysAspAlaLeuHisIleSerGlnIle 658

QY 303 GAAAAAACTATTGGAAAAAGATGAA 329

Db 659 SerLysAsnArgIleGluLysValGlu 667

RESULT 5

US-09-886-055-171

Sequence 171, Application US/09886055

Patent No. US20020132273A1

GENERAL INFORMATION:

APPLICANT: STRYER, LUBERT

APPLICANT: ZOZULYA, SERGEY

TITLE OF INVENTION: RECEPTOR FINGERPRINTING, SENSORY PERCEPTION, AND

TITLE OF INVENTION: BIOSENSORS OF CHEMICAL SENSANTS

FILE REFERENCE: 078003-0277150

CURRENT APPLICATION NUMBER: US/09/886,055

CURRENT FILING DATE: 2001-06-22

PRIOR APPLICATION NUMBER: 60/213,812

PRIOR FILING DATE: 2000-06-22

NUMBER OF SEQ ID NOS: 522

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 171

LENGTH: 312

TYPE: PRT

ORGANISM: Homo sapiens

US-09-886-055-171

Alignment Scores:

Pred. No.: 3.11 Length: 312

Score: 65.00 Matches: 27

Percent Similarity: 37.61% Conservative: 14

Best Local Similarity: 24.77% Mismatches: 50

Query Match: 11.07% Indels: 18

DB: 10 Gaps: 2

US-09-856-221-4 (1-332) x US-09-886-055-171 (1-312)

QY 311 GTTTTTCATCTGTGCAAAATTCAGCATCAAGGTTTCACCTGTCAACAAACAGCATA 252

Db 57 MetTyrPhePheLeuCysAsnLeuSerPheLeuGluIleTyrThrThrValIle 76

QY 251 AGTGCATGCCAGGACCGGGCTTAATGACGCTACAGCTTTATCGTTCAATTCACTAGT 192

Db 77 ProlLysLeuLeuGlyThrPheValValAlaArgThrValIleCysMetSerCysCysLeu 96

QY 191 GTAAGTTTGTTCAGCATCAAAACACAGGAAACAGCAAGTCAATAAACTG----- 141

Db 97 LeuGlnAlaPhePheHisPheValGlyThrThrGluPheLeuIleLeuThrIleMet 116

QY 140 -----ATAATAATAGCCGCAATCGCCACGACGTAGCCAGTTATA 102

Db 117 SerPheAspArgTyrLeuThrIleCysAsnProLeuHisHisProThrIleMetThrSer 136

QY 101 CAACGCTGTATTGCTGAATTTGCTTTGTAGGAGGCGCAGCTGTGCTGTGCTGTGCTG 42

Db 137 LysLeuCysLeu-----GlnLeuAlaLeuSerSerTrpVal 148

QY 41 TTGGTTTCCAGATAGTTTTCGAA 15

Db 149 ValGlyPheThrIleValPheCysGln 157

RESULT 6

US-09-737-068-6

Sequence 6, Application US/09737068

Patent No. US20010020010A1

GENERAL INFORMATION:

APPLICANT: Wallis, Nicola G.

APPLICANT: Zalacain, Madgalenaa

APPLICANT: Throup, John

APPLICANT: Biswas, Sanjoy

TITLE OF INVENTION: Histidine Kinase

NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:

ADDRESSEE: Dechert, Price & Rhoads

STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre

CITY: Philadelphia

STATE: PA

COUNTRY: USA

ZIP: 19103-2793

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: Windows 95

SOFTWARE: FastSeq for Windows Version 2.0b

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/737,068

FILING DATE:

```

; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/081,689
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Falk, Stephen T
; REGISTRATION NUMBER: 36,795
; REFERENCE/DOCKET NUMBER: GM10009
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-994-2488
; TELEFAX: 215-994-2222
; TELEX:
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 218 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-737-068-6

Alignment Scores:
Pred. No.: 3.78 Length: 218
Score: 64.00 Matches: 26
Percent Similarity: 48.94% Conservative: 20
Best Local Similarity: 27.66% Mismatches: 24
Query Match: 10.58% Indels: 6
DB: 10 Gaps: 6

US-09-856-221-4 (1-332) x US-09-737-068-6 (1-218)
Qy 57 GCACAGTGGCC-----TTCCTACAAAGCAATTCAGCAATACAGCGTTGTAACTGG 110
Db 91 AlaserLeuAlaAspGlyLeuGluLysProPheSer-----LeuSerLeu 106
Qy 111 CTAGTGGCGGAGTGGCGCTATTATTATTCAGTTTATCAGTTTGTGCTGCTTCCCTGTGT 170
Db 107 LeuLysValArgValAspAlaIlePheLysArgTyrTyrAspThrGly----- 122
Qy 171 TTGATGCTGAACAACTTACAGATGAATTAACATTAACAGTGAACGTTGCTTCAATGA 230
Db 123 -----ArgIlePheSerTyrLys-----AspThrLysValAspPhe----- 134
Qy 231 CCCGTCCTGGCAGTGGCAGTTCCTGTTTGTAGCAGTGAACCTTGATGCTGAAT 290
Db 135 -----GlusertyrserAlaserLeuAlaGlyGlnGluValProIleasn 149
Qy 291 TTGACACAGTGAAGAA---AAAACTATTGGAAAAAGATGA 329
Db 150 AlalysGluLeuGluIleLeuAspTyrLeuValLysAsnGlu 163

RESULT 7
US-09-801-220-4
; Sequence 4, Application US/09801220
; Patent No. US20020173020A1
; GENERAL INFORMATION:
; APPLICANT: Meyers, Rachel A.
; APPLICANT: Williamson, Mark
; TITLE OF INVENTION: 26886, A NOVEL CARNITINE ACYLTRANSFERASE
; TITLE OF INVENTION: FAMILY MEMBER AND USES THEREFOR
; FILE REFERENCE: 10448-026001
; CURRENT APPLICATION NUMBER: US/09/801,220
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 60/187,456
; PRIOR FILING DATE: 2000-03-07
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows version 4.0
; SEQ ID NO 4
; LENGTH: 698
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: consensus sequence
; US-09-801-220-4

; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/081,689
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Falk, Stephen T
; REGISTRATION NUMBER: 36,795
; REFERENCE/DOCKET NUMBER: GM10009
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-994-2488
; TELEFAX: 215-994-2222
; TELEX:
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 218 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-737-068-6

Alignment Scores:
Pred. No.: 3.78 Length: 218
Score: 64.00 Matches: 26
Percent Similarity: 48.94% Conservative: 20
Best Local Similarity: 27.66% Mismatches: 24
Query Match: 10.58% Indels: 6
DB: 10 Gaps: 6

US-09-856-221-4 (1-332) x US-09-737-068-6 (1-218)
Qy 57 GCACAGTGGCC-----TTCCTACAAAGCAATTCAGCAATACAGCGTTGTAACTGG 110
Db 91 AlaserLeuAlaAspGlyLeuGluLysProPheSer-----LeuSerLeu 106
Qy 111 CTAGTGGCGGAGTGGCGCTATTATTATTCAGTTTATCAGTTTGTGCTGCTTCCCTGTGT 170
Db 107 LeuLysValArgValAspAlaIlePheLysArgTyrTyrAspThrGly----- 122
Qy 171 TTGATGCTGAACAACTTACAGATGAATTAACATTAACAGTGAACGTTGCTTCAATGA 230
Db 123 -----ArgIlePheSerTyrLys-----AspThrLysValAspPhe----- 134
Qy 231 CCCGTCCTGGCAGTGGCAGTTCCTGTTTGTAGCAGTGAACCTTGATGCTGAAT 290
Db 135 -----GlusertyrserAlaserLeuAlaGlyGlnGluValProIleasn 149
Qy 291 TTGACACAGTGAAGAA---AAAACTATTGGAAAAAGATGA 329
Db 150 AlalysGluLeuGluIleLeuAspTyrLeuValLysAsnGlu 163

RESULT 8
US-08-834-666A-2
; Sequence 2, Application US/08834666A
; Patent No. US20020044949A1
; GENERAL INFORMATION:
; APPLICANT: Kleanthous, Harold
; APPLICANT: Kissolo, Ling
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Miller, Charles
; APPLICANT: Al-Garawi, Amal
; TITLE OF INVENTION: 76 kDa Helicobacter Polypeptides and
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESS: Clark & Elbing LLP
; STREET: 176 Federal Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/834,666A
; FILING DATE: 01-APR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,175
; REFERENCE/DOCKET NUMBER: 06132/038001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-428-0200
; TELEFAX: 617-428-7045
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 708 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; FEATURE:
; NAME/KEY: Signal Sequence
; LOCATION: 1...19
; OTHER INFORMATION:
; US-08-834-666A-2

Alignment Scores:
Pred. No.: 5.23 Length: 178
Score: 64.00 Matches: 17
Percent Similarity: 55.56% Conservative: 13
Best Local Similarity: 31.48% Mismatches: 18
Query Match: 10.58% Indels: 6
DB: 9 Gaps: 2

US-09-856-221-4 (1-332) x US-09-801-220-4 (1-698)
Qy 141 CAGTTTATGACTGGCTGTTCCCTGCTTTGATGGCTGAACAACTTACCAGTATGA 200
Db 129 ArgPheLysGluLeuLeuAspAlaSerGluLeuLeuProGluGluLeuAlaLysAsnGlu 148
Qy 201 TTGAACGATAAAGCT-----GTACGCTTCATTAAGCCCGGTGCTGCTGCATGGC 248
Db 149 LysSerAspThrAlaPheLysArgLeuIleArgPheValProSerLeuSerTrpTyrGly 168
Qy 249 ACTTATGCTGGTTGTAGCAGCTGAACCTTCATGCTGAAT 290
Db 169 AlaTyr-----LeuLeuGlyGlyGlnProLeuCysMetAsn 180

RESULT 9
US-08-834-666A-2
; Sequence 2, Application US/08834666A
; Patent No. US20020044949A1
; GENERAL INFORMATION:
; APPLICANT: Kleanthous, Harold
; APPLICANT: Kissolo, Ling
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Miller, Charles
; APPLICANT: Al-Garawi, Amal
; TITLE OF INVENTION: 76 kDa Helicobacter Polypeptides and
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESS: Clark & Elbing LLP
; STREET: 176 Federal Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/834,666A
; FILING DATE: 01-APR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,175
; REFERENCE/DOCKET NUMBER: 06132/038001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-428-0200
; TELEFAX: 617-428-7045
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 708 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; FEATURE:
; NAME/KEY: Signal Sequence
; LOCATION: 1...19
; OTHER INFORMATION:
; US-08-834-666A-2
```

Alignment Scores:  
Pred. No.: 6.08 Length: 708  
Score: 63.50 Matches: 26  
Percent Similarity: 35.79% Conservative: 8  
Best Local Similarity: 27.37% Mismatches: 32  
Query Match: 10.50% Indels: 29  
DB: 8 Gaps: 3

US-09-856-221-4 (1-332) x US-08-834-666A-2 (1-708)

QY 66 GCCTTCTACAAAGCAAAATTCAGCATACAGCGTGTATACACTGGCTACGGGCGCATG 125  
Db 563 AlaPheIleLysSerSerPheAsnSerAlaSerAspValTrpThrTyrGlyPheGly 582  
QY 126 GCGCGTATTATATATCAGTTTATGACTTGGCTGTTCCCTGTTGTTGATGCGTGACAA 185  
Db 583 AlaAspAlaLeuTyrAsnPhe 589  
QY 186 ACTTACAGTATGAATGAACGATGAAGCTGACGCTTCATTAAGCCCGTGGCTGGCAT 245  
Db 590 -----IleAsnAspLysAlaThrAsnPheLeu-----GlyLysAsnAsn 602  
QY 246 GGCACATTATGCTGGTGTGTAGCAGCTGAACCTTG----- 281  
Db 603 LysLeuSerValGlyLeuPheGlyIleAlaLeuAlaGlyThrSerTrpLeuAsnSer 622  
QY 282 ---ATGCTGAATTTGGCAGACAGATGAGAAAAAACTATTGGAAAAA 323  
Db 623 GluTyrValAsnLeuAlaThrValAsnAsnValTyrAsnAlaLys 637

RESULT 9  
US-09-815-242-13932  
; Sequence 13932, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13932  
; LENGTH: 489  
; TYPE: PRT  
; ORGANISM: Salmonella typhi  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (1)...(489)  
; OTHER INFORMATION: Xaa = Any Amino Acid





```
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
; SEQ ID NO 37729
; LENGTH: 47
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC006344.2
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 0.93
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.93
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.87
; OTHER INFORMATION: EXPRESSED IN HEPA, SIGNAL = 0.81
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.92
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.86
; OTHER INFORMATION: EST_HUMAN HIT: BE466313.1, EVALUATE 1.00e-11
; OTHER INFORMATION: EST_HUMAN HIT: BE466313.1, EVALUATE 1.00e-11
US-09-856-221-4 (1-332) x US-09-864-761-37729 (1-47)
Alignment Scores:
Pred. No.: 6 Length: 47
Score: 61.00 Matches: 18
Percent Similarity: 52.50% Conservative: 3
Best Local Similarity: 45.00% Mismatches: 5
Query Match: 10.08% Indels: 14
DB: 10 Gaps: 4
US-09-856-221-4 (1-332) x US-09-864-761-37729 (1-47)
QY 165 CTGTGTTTGGCTGACAACTACAGTATGATTCAGCGATAAAGCTGTACGCTC 224
Db 16 LeuCysPheLeu-----ThrTyrLysTrpGluLeuAsnAspLys----- 28
QY 225 ATTAAGCCGGTGCTCGG---CATGGC-----ACATTATGCTGGTTTGTAGCAGGT 272
Db 29 -----AsnLysTrpThrHisGlyGlyAsnAsnThrTyrTrpGlyLeuLeuGluGly 45
RESULT 13
US-09-764-864-996
; Sequence 996, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ23
; CURRENT APPLICATION NUMBER: US/09/764,864
; PRIOR FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 996
; LENGTH: 359
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-864-996
Alignment Scores:
Pred. No.: 10.6 Length: 359
Score: 61.00 Matches: 19
Percent Similarity: 43.86% Conservative: 6
Best Local Similarity: 33.33% Mismatches: 14
Query Match: 10.39% Indels: 18
DB: 10 Gaps: 3
US-09-856-221-4 (1-332) x US-09-764-864-996 (1-359)
QY 291 AATTGAGCATCAAGGTTTCACCTGCTTAACAAACAGCATAAGTGCAGCGACCGG 232
Db 87 AsnSerSerSer-----LeuLeuAsnHisHisLysValHisAla----- 99
QY 231 GCTTAATGAAGCGTACAGCTTTATCGTTCATTCATCTACCTGGTAAGTTGTTTCAGCCATCA 172
Db 100 -----GlyLysGlnProTyrArgCysAlaIleCysGlyLysPheLeuLysLysHis 116
QY 171 AACACA-----GGGAAACAGCAAGTCATAAA 145
Db 117 SerThrPheIleAsnHisGlnArgIleHisSerArgGluLysProHisLys 133
RESULT 14
US-09-815-242-12036
; Sequence 12036, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlson, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12036
; LENGTH: 487
; TYPE: PRT
```

; ORGANISM: Pseudomonas aeruginosa  
US-09-815-242-12036

Alignment Scores: 11.5 Length: 487  
Pred. No.: 61.00 Matches: 28  
Score: 36.07% Conservative: 16  
Best Local Similarity: 22.95% Mismatches: 31  
Query Match: 10.08% Indels: 47  
DB: 10 Gaps: 6

US-09-856-221-4 (1-332) x US-09-815-242-12036 (1-487)

Qy 63 CTGGCTTCTCAAGCAAAATTGACGCTTGTATTAACCTGCTACGTGGCGGA 122  
Db 357 LeuCySPheLeuThrSerAlaPheGlyAspSerThValTyThrTriPleuLeuAsnThr 376  
Qy 123 -----TTGGCGGCTATTATTATCAGTTT--- 146  
Db 377 SerGlyMetCysGlyPheIleAlaTrpLeuGlyIleAlaIleSerHisTyArgPheArg 396  
Qy 147 -----TATGACTTG----- 155  
Db 397 LysGlyTyTrpLeuAlaGlnGlyArgLeuGluAspLeuProTyArgAlaLysLeuPhe 416  
Qy 156 -----GCTGTTTCCTGTGTGTGTGTGTGCTGAA-----CAAACTTAC 191  
Db 417 ProPheGlyProLeuPheAlaPheAlaLeuCysMetValIleThrLeuGlyGlnAsnTy 436  
Qy 192 CAGTATGAAATTGAAGCAATAAGCTGTAGCTTCATTAAAGCCCGGTCCCTGGCAGTGC 251  
Db 437 GlnAlaLeuValGlyGluArgIle-----AspTrpIleGlyLeu 449  
Qy 252 TATGCTGGTGTGTAGCAGGTGAAACCTTGATGCTGAATTTGG-----CACAGATGAA 305  
Db 450 LeuAlaThrTyIleSerLeuProLeuPheLeuAla-IleTrpLeuGlyTyArgTrpLy 469  
Qy 306 AAAA 309  
Db 469 sLys 470

RESULT 15

US-09-861-451A-80  
; Sequence 80, Application US/09861451A  
; Patent No. US20020068289A1

; GENERAL INFORMATION:  
; APPLICANT: Commonwealth Scientific & Industrial Research Orga

; TITLE OF INVENTION: Methods of Identifying Antigen Gene Sequences  
; FILE REFERENCE: FF34033/01

; CURRENT APPLICATION NUMBER: US/09/861,451A  
; PRIOR FILING DATE: 2001-05-21

; PRIOR APPLICATION NUMBER: PP7273  
; PRIOR FILING DATE: 1998-11-20

; NUMBER OF SEQ ID NOS: 84  
; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 80  
; LENGTH: 163  
; TYPE: PRT

; ORGANISM: Artificial Sequence  
; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence:Deduced protein  
; OTHER INFORMATION: Sequence from clone PAD766

US-09-861-451A-80

Alignment Scores: 9.83 Length: 163  
Pred. No.: 60.50 Matches: 14  
Score: 52.17% Conservative: 10  
Percent Similarity: 30.43% Mismatches: 21  
Best Local Similarity: 10.00% Indels: 1  
Query Match: 10 Gaps: 1  
DB: 1

US-09-856-221-4 (1-332) x US-09-861-451A-80 (1-163)

Qy 189 TACAGTATGAATTGACGATAAAAGCTGTACGCTTCAATTAAAGCCGGTGCCTGGCATGGC 248  
Db 45 TyrLysTyTrpLysPheAsnAspLysIleHisLysPhe---GluProThrGluTyTrLys 63  
Qy 249 ACTTATGCTGCTTGTAGCAGGTGAAACCTTGATGCTGAATTTGGCACACATGGAAAAA 308  
Db 64 AsnThrAlaAsnPheSerGlnGlyGlyLeuTySerAlaAsnLeuLeuGluLeuLys 83  
Qy 309 AACTATTGGAAGAAAGAT 326  
Db 84 GluIleLysLysGlnAsp 89

Search completed: January 15, 2003, 15:58:44  
Job time : 10.2284 secs





## Alignment Scores:

Pred. No.: 4.07 Length: 598  
Score: 70.50 Matches: 37  
Percent Similarity: 35.8% Conservative: 15  
Best Local Similarity: 25.52% Mismatches: 35  
Query Match: 11.65% Indels: 58  
DB: 2 Gaps: 6

US-09-856-221-4 (1-332) x D71365 (1-598)

QY 53 TCAGGCACAGCTGGCTCCCTACAAAGCAATTCAGCAATACAGCCTTGTATTAACCTGGCT 112  
Db 19 SerGlyTyrAlaGlyValLeuThrProGlnVal-SerGlyThrAlaGlnLeuGlnTrp-- 37  
QY 113 ACCTGGCGATGGCGGCTATTATTATTCAG----- 143  
Db 38 -----GlyIleAlaPheGlnLysAsnProArgThrGlyProGlyLysHI 52  
QY 144 -----TTTTAGACTGGCTGGTTCCCTGTGTTGATGGCTGAACA 184  
Db 52 sThrHisGlyPheArgThrThrAsnSerLeuThrIleSerLeuProLeuValSerLysHI 72  
QY 185 AACTTAC-----CAGTATGAATTGAACGA 208  
Db 72 sThrHisThrArgArgGlyGluAlaArgSerGlyValTrpAlaGlnLeuLysAs 92  
QY 209 TAAAGCTGACGC-----TTTCATTAAAGCCCGG 235  
Db 92 pLeuAlaValGluLeuAlaSerSerLysSerThrAlaLeuSerPheThrLysProTh 112  
QY 236 TGCC-----TGGCATGGCATTGCTGGTTGTTAGCAGGTGA 274  
Db 112 rAlaSerPheGlnAlaThrLeuHisCysTyrGlyAlaTyrLeuThrValGlyThrSerPr 132  
QY 275 AACCTTGATGCTGAATTGGCAGATGGAAAA-----AACTATTGGA 319  
Db 132 oSerCysValValAsnPheAlaGlnLeuTrpLysProPheValThrArgAlaTyrSerGl 152  
QY 320 AAAAGTGAACGG 332  
Db 152 ulysAspThrArg 156

RESULT 5  
F82207  
ATP-dependent helicase hRPA VCI382 [imported] - Vibrio cholerae (strain N16961 serogroup  
C;Species: Vibrio cholerae  
C;Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001  
C;Accession: F82207  
R;Heidelber, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;  
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, F.  
I, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
Nature 406, 477-483, 2000  
A;Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.  
A;Reference number: A82035; MUID:20406833; PMID:10952301  
A;Accession: F82207  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-1309 <HEI>  
A;Cross-references: GB:AE004217; GB:AE003852; NID:9655866; PID:AAF94540.1; GSPDB:GN001  
A;Experimental source: serogroup O1; strain N16961; biotype El Tor  
C;Genetics:  
A;Gene: VCI382  
A;Map position: 1

Alignment Scores:  
Pred. No.: 4.31 Length: 1309  
Score: 70.50 Matches: 23  
Percent Similarity: 43.43% Conservative: 20  
Best Local Similarity: 23.23% Mismatches: 31  
Query Match: 11.65% Indels: 25  
DB: 2 Gaps: 4

US-09-856-221-4 (1-332) x F82207 (1-1309)

QY 18 CAAAAAACTATCTGGAACCCACAGGCACAACTCAGGCACAGCTGGCTTCTCTACAA 77  
Db 579 GlnGlnLysAlaLeuThrSerAsnGlnPheArgGlnCysLysLeuAspTyrLeu--- 597  
QY 78 AGCAAAATTCAGCAATACAGCGTTGTATTAACCTGGCTACGTGGCGCA---TTGGCGGCTATT 134  
Db 598 -----AsnTyrLeuArgValArgGluTrpGlnAspVal 608  
QY 135 TATTATCAGTTTTATGACTTTGGCTGTTCCTGTGTTTGGCTGGAACAACTTACCAG 194  
Db 609 TyrThrGlnLeuHis-----GlnSerThrArgGluMetGly 620  
QY 195 TATGAATTCAGCAATGAAGCTGACGCTTCATTAAGCCCGGTGGCTGGCATGGCACTTAT 254  
Db 621 PheLysLeuAsnAsp-----GluProGlySerTyrTyrHisAlaValHis 634  
QY 255 GCTGGTTGTTAGCAGCTGAAACCTTTGATGCTGAATTTGGCACAGATGGAAAAAAC 311  
Db 635 SerAlaIleLeuValGlyLeuLeuSerHisIleGlyMetLysAspGlnGluLysAsn 653

RESULT 6  
A21185  
hypoetical protein - Trypanosoma cruzi (fragment)  
C;Species: Trypanosoma cruzi  
C;Date: 07-Sep-1990 #sequence\_revision 13-Mar-1997 #text\_change 13-Mar-1997  
C;Accession: A21185  
R;Gonzalez, A.; Prediger, E.; Huecas, M.E.; Nogueira, N.; Lizardi, P.M.  
Proc. Natl. Acad. Sci. U.S.A. 81, 3356-3360, 1984  
A;Title: Minichromosome repetitive DNA in Trypanosoma cruzi: its use in a high-sensi  
A;Reference number: A21185; MUID:84221958; PMID:6427769  
A;Accession: A21185  
A;Molecule type: DNA  
A;Residues: 1-194 <GON>  
A;Note: the authors translated the codon GAG for residue 15 as Gln

Alignment Scores:  
Pred. No.: 4.33 Length: 194  
Score: 70.00 Matches: 19  
Percent Similarity: 50.94% Conservative: 8  
Best Local Similarity: 35.85% Mismatches: 24  
Query Match: 11.57% Indels: 2  
DB: 2 Gaps: 2

US-09-856-221-4 (1-332) x A21185 (1-194)

QY 16 TGCACAAAAATATCTGGAACCCACAGGCACAACTCAGGCACAGCTGGCTTCTCTAC 75  
Db 45 CysValPheThrHisThrProAsnAsnProGluLeuSerAlaAlaTrpArgAsnPhe 64  
QY 76 AAAGCAATTCAGCAATACAGCGTTGTATACTGGC---TACGTGGCGGATTTGGCGGCTA 132  
Db 65 ---AlaSerSerCysProGlnGlyCysCysThrArgLeuIleValPheGluTrpLeuLeu 83  
QY 133 TTTATTATCAGTTTATGACTTGGCTGTTCCTCGTGT 171  
Db 84 HisHisThrLeuTrpSerLysPheLeuPheProIleVal 96

RESULT 7  
E81828  
probable periplasmic type I secretion system protein NMA1994 [imported] - Neisseria m  
C;Species: Neisseria meningitidis  
C;Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 02-Feb-2001  
C;Accession: E81828  
R;Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Mo  
; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandre  
Nature 404, 502-506, 2000  
A;Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 22491  
A;Reference number: A81775; MUID:20222556; PMID:10761919  
A;Accession: E81828  
A;Status: preliminary  
A;Molecule type: DNA

A:Residues: 1-467 <PAR>  
A:Cross-references: GB:AL157959; NID:g7380371; PIDN:CAB85213.1; PID:g738062  
A:Experimental source: serogroup A, strain Z2491  
C:Genetics:  
A:Gene: NMA1992; NMA1994

Alignment Scores:  
Pred. No.: 5.33 Length: 467  
Score: 69.50 Matches: 30  
Percent Similarity: 42.99% Conservative: 16  
Best Local Similarity: 28.04% Mismatches: 38  
Query Match: 11.49% Indels: 23  
DB: 2 Gaps: 5

US-09-856-221-4 (1-332) x E81828 (1-467)

QY 33 GARACCCACAGCAGCACTACAGCAGCTGCCTCTCTACAAAGCAATTCAGCAAT 92  
Db 176 AspThrValAlaAlaHisAlaAlaGluLysGluAlaThrAlaGlnGlnValargGlnAla 195  
QY 93 ACAGCGTGTATACCTGCTAGCTGGGGGATTGGCGCTATTATTATCAG----- 143  
Db 196 GlnAlaLeuPheAsn-----LysGlyAlaAlaThrAlaLeuAspIleHisGluAlaLys 213  
QY 144 ---TTTATCAG-----TTGGCTGTTCCCTGTGTGTGATGCTGAA---CAAACTTAC 191  
Db 214 AlaGlyPyrAspAsnAlaLeuAlaGlnGluIleAlaValLeuAlaGluLysGlnThrTyr 233  
QY 192 CAGTATGAATTAACGATTAAGCTCTACGCTTCATTAAAGCCCGTGCCTGGCATGCACT 251  
Db 234 GluAsnGlnLeuAsnAsp----- 239

QY 252 TATGCTGGTGTGTACAGCTGAACCTTGATGCTGATTTGGCAGATGGAAAAAAC 311  
Db 240 TyrThrGlyLeuAspSerLysGlnLeuAlaIleAspThrAlaAsnLeuLeuAlaArg 259  
QY 312 TATTGGAAAAAGATGAACGG 332  
Db 260 TyrLeuProLysLeuGluArg 266

RESULT 8  
NDECKR  
type I site-specific deoxyribonuclease (EC 3.1.21.3) EcoK chain R - Escherichia coli (str  
N:Alternate names: type I restriction enzyme EcoK R chain; type I restriction-modificati  
C:Species: Escherichia coli  
C:Date: 31-Mar-1990 #sequence\_revision 26-Sep-1997 #text\_change 01-Mar-2002  
C:Accession: S56576; H65249; S18776; A30375; Q00648  
R:Burland, V.; Plunkett III, G.; Sofia, H.J.; Daniels, D.L.; Blattner, F.R.  
Nucleic Acids Res. 23, 2105-2119, 1995  
A:Title: Analysis of the Escherichia coli genome VI: DNA sequence of the region from 92.  
A:Reference number: S56314; MUID:95334362; PMID:7610040  
A:Accession: S56576

A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-1188 <RUP>  
A:Cross-references: EMBL:U14003; NID:gl263172; PIDN:AAA97247.1; PID:g537192  
A:Note: The nucleotide sequence was submitted to the EMBL Data Library, August 1994  
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co  
.A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A:Title: The complete genome sequence of Escherichia coli K-12.  
A:Reference number: A64720; MUID:97426617; PMID:9278503  
A:Accession: H65249  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-1188 <BLAT>

A:Cross-references: GB:AE0000505; GB:U000096; NID:g2367375; PIDN:AACT7306.1; PID:g1790809;  
A:Experimental source: strain K-12, substrain MG1655  
R:Waite-Rees, P.A.; Keating, C.J.; Moran, L.S.; Slatko, B.E.; Hornstra, L.J.; Benner, J.  
J. Bacteriol. 173, 5207-5219, 1991  
A:Title: Characterization and expression of the Escherichia coli mrr restriction system.  
A:Reference number: A40368; MUID:91317743; PMID:1650347  
A:Accession: S18776

A:Status: translation not shown  
A:Molecule type: DNA  
A:Residues: 1-27 <WAI>  
A:Cross-references: EMBL:X54198  
R:Loenen, W.A.M.; Daniel, A.S.; Braymer, H.D.; Murray, N.E.  
J. Mol. Biol. 198, 159-170, 1987  
A:Title: Organization and sequence of the hsd genes of Escherichia coli K-12.  
A:Reference number: A30375; MUID:88118919; PMID:3323532  
A:Accession: A30375

A:Molecule type: DNA  
A:Residues: 1-628, 'ECGNKSHRHPGATYCADFRAGLPLYLPRGYRRFSDRPSAYSDHHPQAGGLSLQRRAGRAH  
A:Cross-references: GB:X06545; NID:g41751; PIDN:CAA29791.1; PID:941752  
C:Comment: This is one of three components (S, R, and M chains) of type I site-specif  
ctors: It is the site-specificity determinant for the catalytic activities of the enz  
C:Genetics:  
A:Gene: hsdR  
A:Map position: 99 min  
C:Superfamily: type I site-specific deoxyribonuclease EcoK chain R  
C:Keywords: Atp; DNA binding; hydrolase; nucleotide binding; P-loop; restriction mod  
F:489-496/Region: nucleotide-binding motif A (P-loop)  
F:588-593/Region: nucleotide-binding motif B  
F:592-595/Region: DEAH motif

Alignment Scores:  
Pred. No.: 5.7 Length: 1188  
Score: 69.50 Matches: 30  
Percent Similarity: 37.82% Conservative: 15  
Best Local Similarity: 25.21% Mismatches: 33  
Query Match: 11.49% Indels: 41  
DB: 1 Gaps: 5

US-09-856-221-4 (1-332) x NDECKR (1-1188)

QY 21 AAAAATCTATCGGAACCCAA-----CAGGCACAACTCAGGCACAGCTGCCTTC 71  
Db 169 LysGlnGlnLeuGlnGlnValArgGluLysAlaGlnThrGlnAlaGluValGluAla 188  
QY 72 CTACAAAGCAAAATTCAGCAATACAGCGTGTGTATTAAGTGGCTA----- 113  
Db 189 GlnGlnGlnLysLeu-----ValAlaLeuAsnGlyTyrIleAlaLeuGluGluLys 206  
QY 114 -----CGTGGCGGATTTGGCGGCTATTATTATCAGCTTTAT 149  
Db 207 GlnGlnGluThrGluAlaGlnThrGlnAlaArgLeuAlaLeuGluAlaGlnLeuAla 226  
QY 150 GACTTGGCTGTTCCCTGCTGTTGATGGCTGAACAA-----ACTTACCAGTATGAA 200  
Db 227 GluLysAsnAlaGluLeuAlaLysGlnThrGluGlnGluArgLysAlaTyrHisLysGlu 246  
QY 201 TTGACAGTAAAGTGTACGCTTCATTACCCCGCTGCCTGGCATGCGACTATGCTGGT 260  
Db 247 IleThrAspGlnAlaIleLys----- 253  
QY 261 TTGTTAGCAGGTGAACCTTGATGCTGAATTTGGCAGATGGAAAAAACTATTGT 317  
Db 254 -----ArgThrLeuAsnLeuSerGluGluGluSerArgPheLeu 266

RESULT 9  
A43336  
microtubule-vesicle linker CLIP-170 - human  
C:Species: Homo sapiens (man)  
C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 05-Nov-1999  
C:Accession: A43336  
R:Pierrer, P.; Scheel, J.; Rickard, J.E.; Kreis, T.E.  
Cell 70, 887-900, 1992

A:Title: CLIP-170 links endocytic vesicles to microtubules.  
A:Reference number: A43336; MUID:92405160; PMID:1356075  
A:Accession: A43336  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-1392 <PLE>  
A:Cross-references: GB:M97501; NID:g180621; PIDN:AAA35693.1; PID:g180622



Db 196 aSerProHisPheAsnProGlnSerSerGluLysSerPheAsnGlnAspTyrAsnThrVa 216  
 QY 235 -----GTCCTGGCATGGCACTATGCTGGTTGTAGCAGGTGAACCTTGATGCTG 287  
 Db 216 1AspGluLeuPro----- 220  
 QY 288 AATTGGCAGCATGGAAGAACTATTGGAAG 324  
 Db 221 ----TrrTyrLysTrpLysGlyHisValTrpAsnLys 231  
 RESULT 12  
 S75950  
 hypothetical protein - *Synechocystis* sp. (strain PCC 6803)  
 C:Species: *Synechocystis* sp.  
 A:Variety: PCC 6803  
 C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 08-Oct-1999  
 C:Accession: S75950  
 R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;  
 O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda  
 DNA Res. 3, 109-136, 1996  
 A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis*  
 s.  
 A:Reference number: S74322; MUID:97061201; PMID:8905231  
 A:Accession: S75950  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-258 <KAN>  
 A:Cross-references: EMBL:D64006; GB:AB001339; NID:q1001291; PIDN:RAAL0797.1; PID:d101144  
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996  
 Alignment Scores:  
 Pred. No.: 6.79 Length: 258  
 Score: 68.50 Matches: 23  
 Percent Similarity: 37.50% Conservative: 13  
 Best Local Similarity: 23.98% Mismatches: 29  
 Query Match: 11.32% Indels: 31  
 Gaps: 5  
 DB:  
 US-09-856-221-4 (1-332) x S75950 (1-258)  
 QY 57 GCACAGTGGCTTCTTACAGCAATACAGCGTTGTATACACTGGCTACGT 116  
 Db 186 AlaGlnAlaThrHisLeuSerAlaSerPheAlaGlnThrLysAlaPheAsnTrpLeuGln 205  
 QY 117 GGGCGATTGGCGCTATTATTATCATGTTTATGACTTGGCTGGTTCCTGCTTTGATG 176  
 Db 206 AsnAsnAlaAla----- 209  
 QY 177 GCTGAACAACCTTACCAGTATGANTG-----AACGATAAGCTGTACGC 221  
 Db 210 -----LysTyrSerPheGluLeuSerPheProProAspAsnProGlnGlyIleAla 226  
 QY 222 TTCATTAAAGCCCGTGGCTGGCATGTCACATTATGCTGTTTGTAGCAGGTGAACCTTG 281  
 Db 227 Tyr---GluPro-----TrpHisTrpArgTyr-----ValGlyAspArgGln 239  
 QY 282 ATGCTGAATTGGCAGACAGATGGAAGAACTATTGGAAAAAGATGAA 329  
 Db 240 SerLeuGluLeuPheTyrLysAlaArgAsnLeuProGlnLysAsnGlu 255  
 RESULT 13  
 T33215  
 hypothetical protein T07H8.5 - *Caenorhabditis elegans*  
 C:Species: *Caenorhabditis elegans*  
 C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
 C:Accession: T33215  
 R:Gattung, S.; Maggi, L.  
 submitted to the EMBL Data Library, May 1998  
 A:Description: The sequence of *C. elegans* cosmid T07H8.  
 A:Reference number: Z21303  
 A:Accession: T33215  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA

A:Residues: 1-312 <GAT>  
 A:Cross-references: EMBL:AF067945; PIDN:AAC17679.1; GSPDB:GN00023; CESP:T07H8.5  
 A:Experimental source: strain Bristol N2; clone T07H8  
 C:Genetics:  
 A:Gene: CESP:T07H8.5  
 A:Map position: 5  
 A:Introns: 48/3; 126/3; 236/3; 282/3  
 Alignment Scores:  
 Pred. No.: 6.88 Length: 312  
 Score: 68.50 Matches: 35  
 Percent Similarity: 41.67% Conservative: 15  
 Best Local Similarity: 29.17% Mismatches: 30  
 Query Match: 11.32% Indels: 40  
 Gaps: 5  
 DB:  
 US-09-856-221-4 (1-332) x T33215 (1-312)  
 QY 9 GCAGTATTGCAAAAACTATCTG---GAACCCCAACAGGCACAACTCAGGCACAGCTG 65  
 Db 155 AlaTyrLeuGluTyrAsnProLeuSerAspThrTyrValAlaArgThrGlnAlaAspLeu 174  
 QY 66 GCCTTCCTACAA----- 77  
 Db 175 ProPheLeuTyrSerPheIleLeuValTrpMetValValThrValLeuLeuSerIleIle 194  
 QY 78 -----AGCAATTCAGCAATACAGCGTTGTATACCTGG 110  
 Db 195 AlaAsnIleIleCysTrrPheLysIleSerLysTyrSerLysAlaAlaArgGln----- 212  
 QY 111 CTACGTGGCGGATGGCGGCTATTATTATCAGTTTATGACTGGCTGTTCCCTCTG-- 168  
 Db 213 ---Gln-SerAspTyrArgLeuPheLeuValSerPheValThrPheValIleAsnCysG1 231  
 QY 169 -----GTTTGTATGGCTGAACAACTTACAGTATGAATTGACGATATAAGCTGTA 218  
 Db 231 yValPheSerIleAlaMetLeuAsnLysIleSerAlaAspIleAspProSerLysLeuLe 251  
 QY 219 CGCTTCA-----TTAGCCGGCTGCTGCATGGCAGCTTATGCTG 258  
 Db 251 uLeuSerArgIleAlaGlnLeuLeuSerPropheAlaAsnAspLeuLeuSerLeu 270  
 RESULT 14  
 AH0159  
 Probable lysine-specific permease *lypS* [imported] - *Yersinia pestis* (strain CO92)  
 C:Species: *Yersinia pestis*  
 C:Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 09-Nov-2001  
 C:Accession: AH0159  
 R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Hoiden, M.T.G.; Prentice, M.  
 deno-Parraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.  
 il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell  
 Nature 413, 523-527, 2001  
 A:Title: Genome sequence of *Yersinia pestis*, the causative agent of plague.  
 A:Reference number: AH00001; MUID:21470413; PMID:11586360  
 A:Accession: AH0159  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-503 <KUR>  
 A:Cross-references: GB:AL590842; PIDN:CAC90139.1; PID:gl15979359; GSPDB:GN00175  
 C:Genetics:  
 A:Gene: *lypS*  
 C:Superfamily: arginine permease  
 Alignment Scores:  
 Pred. No.: 8.22 Length: 503  
 Score: 68.00 Matches: 35  
 Percent Similarity: 35.07% Conservative: 12  
 Best Local Similarity: 26.12% Mismatches: 29  
 Query Match: 11.24% Indels: 58  
 Gaps: 8  
 DB:  
 US-09-856-221-4 (1-332) x AH0159 (1-503)



Search completed: January 15, 2003, 15:56:29  
Job time : 20.3002 secs

```
QY 57 GCACAGCTGGCTTCTACAAAGCAATTCAGCAATACAGCGTTGTATTAAGTGGCTA--- 113
Db 360 AlaglyLeuCysPheLeuSerSerMetPheGlyAsnGlnThrValTyrLeuTrpLeuLeu 379
QY 113 ----- 113
Db 380 AsnThrSerGlyMetThrGlyPheIleAlaTrpLeuGlyIleAlaIleSerHisTyrArg 399
QY 114 -----CGTGGCGCA---TTGGCGGCTATTTATTATCAG----- 143
Db 400 PheArgArgGlyTyrMetMetGlnGlyArgAspLeuAsnAspLeuProTyrGlnSerGly 419
QY 144 TTTTATGACTTG-----GCTGTTTCCCTGCTGTTTGTATGGCTGAA-----CAA 185
Db 420 PhePheProLeuGlyProIlePheAlaPheValLeuCysLeuIleIleThrLeuGlyGln 439
QY 186 ACTTACCAGTATGANTTGAACGATAAGCTGTACCTTCATTAAAGCCCGGTGGCTGGCAT 245
Db 440 AsnTyrGlnAlaPheLeuGlnAspArgIle-----AspTrpTyr 452
QY 246 GGC-----ACTTATGCTGCTTTG----- 263
Db 453 GlyValThrAlaThrTyrIleGlyIleProLeuPheLeuValIleTrpPheGlyTyrLys 472
QY 264 TTACAGAGTGAACCTTGATGCTGTAATTTGGCACAGATGAA 305
Db 473 LeuSerArgGlyThrArgValValLysTyrGlnGluMetGlu 486

RESULT 15
E89901
polyribonucleotide nucleotidyltransferase [imported] - Staphylococcus aureus (strain N31)
C:Species: Staphylococcus aureus
C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C:Accession: E89901
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc
ma, A.; Mizutani-Tajima, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A:Reference number: A89758; PMID:21311952; PMID:11418146
A:Accession: E89901
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-698 <CUR>
A:Cross-references: GB:BA000018; PID:g13701074; PIDN:BA842369.1; GSPDB:GN00149
A:Experimental source: strain N315
C:Genetics:
A:Gene: pnpA
C:Superfamily: polyribonucleotide nucleotidyltransferase alpha chain

Alignment Scores:
Pred. No.: 8.42 Length: 698
Score: 68.00 Matches: 17
Percent Similarity: 57.14% Conservative: 11
Best Local Similarity: 34.69% Mismatches: 13
Query Match: 11.24% Indels: 8
DB: 2 Gaps: 2

US-09-856-221-4 (1-332) x E89901 (1-698)
QY 183 CAACCTTACCAGTATGAATGAACGATAAAGCTGTACGCTTCATTAAAGCCCGGTGGCTGG 242
Db 627 GlnThrTyrGln-----AlaThrValLysArgIleGluLys----- 638
QY 243 CATGGCACTTATGCTGTTTGTATGAGTGAACCTTGATGCTGAATTTGGCACAGATG 302
Db 639 TyrGlyAlaPheValGlyLeuPheProGlyLysAspAlaLeuLeuHisIleSerGlnIle 658
QY 303 GAAAGAACTATTGGAAGATGAA 329
Db 659 SerLysAsnArgIleGluLysValGlu 667
```

